On Kant, Arpaly and Practical Rationality

Dissertation

Presented in Partial Fulfillment of the Requirement for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Andrew N. Choi, B.A.

Graduate Program in Philosophy

The Ohio State University

2010

Dissertation Committee:

Timothy Schroeder, Advisor

Sigrún Svavarsdóttir

Piers Norris Turner
Abstract

In my dissertation, I develop and defend a new version of the Kantian theory of action and practical rationality. According to the Kantian theory of action, action is behavior that is genuinely guided by a series of representations which reflect the agent's understanding of what he is doing – namely, judgments about reasons for action, best judgments and choices. Moreover, one significant claim of the Kantian theory of practical rationality is the claim that rational action is action that accords with one's best judgment, and irrational action is action that goes against one's best judgment - where acting in accordance with one's best judgment is taken to be a necessary (and not sufficient) condition of rational action, whereas acting against one's best judgment is taken to be a sufficient condition of irrational action.

While these claims about action and practical rationality have been taken for granted by Kantians, they have recently become the subject of much criticism by Nomy Arpaly. In a well-known paper, Arpaly presents a character named Sam, an agent who allegedly acts rationally even though he acts against his best judgment. Given that Sam acts rationally, Arpaly concludes that best judgment is irrelevant to a proper account of rational and irrational action. In light of this, the aforementioned Kantian claim about rational and irrational action, a claim which is endorsed by many others as well, looks to be problematic. Now, while the main target of Arpaly's argument is Kant's theory of
practical rationality, and more specifically, its claims about rational and irrational action, Arpaly's argument has repercussions for Kant's theory of action as well. As it turns out, the Kantian theory of action makes additional claims that make it quite difficult for the Kantian to give an adequate account of Sam.

The key point of contention concerns an additional claim that Kantians commit themselves to, regarding how one accounts for the agency in action. As I read Kant, the agency in action is manifested in the connection that the representations which guide the agent's action have to the awareness of the agent. For Kant, to say that an action is something that the agent does is to say that the action is guided by the agent's understanding. Moreover, to say that an action is guided by the agent's understanding is to say that the action is guided by representations which are conscious states of the agent's mind. These claims are problematic because they place a heavy psychological burden on agents like Sam. Given that Sam is an agent who acts rationally, the Kantian must claim that a very specific set of mental states be conscious states of Sam's mind. However, given the way in which Arpaly has characterized Sam, this is simply not the case. In light of the difficulties posed by the case of Sam, Arpaly thinks we have good reason to scrap the Kantian project in its entirety, in favor of her own theory of action and practical rationality.

Contrary to Arpaly, I am less pessimistic about the Kantian enterprise. I develop a theory of action and practical rationality that coopts much of the Kantian framework, but revises the claims that Kantians make about the connection between agency, understanding and consciousness. In contrast with the traditional Kantian, I
argue that there is good reason to think that unconscious mental states can be manifestations of the agent's understanding in action, albeit in a limited set of cases. Building on the work of Kant, Barbara Herman and Nomy Arpaly, I then explain how agents can, in these cases, form best judgments and choices in such a way that they are unconscious.

In light of the changes, the new theory accommodates the case of Sam quite well – in fact it shows that Sam can be characterized as an agent who acts rationally and acts *in accordance* with his best judgment. The end result is a theory of action and practical rationality that provides a new alternative for the Kantian – an alternative which resists Arpaly's push to dispense with the Kantian framework. Furthermore, it preserves a central tenet that Kantians and many others hold dear – that rational action is action that accords with one's best judgment, and irrational action is action that goes against one's best judgment.
Dedicated to the students at The Ohio State University
Acknowledgments

I would not have been able to finish this dissertation without help from my committee members, Timothy Schoeder (advisor), Sigrún Svavarsdóttir, Piers Norris Turner and former committee member Dan Farrell.
Vita

June 1998......................................................Ames High School

June 2002......................................................B.A. Princeton University

Fields of Study

Major Field: Philosophy
# Table of Contents

Abstract........................................................................................................................................ ii  
Dedication..................................................................................................................................... v  
Acknowledgments.................................................................................................................... vi  
Vita............................................................................................................................................... vii  
Chapter 1: An Introduction to the Kantian Framework............................................................. 1  
Chapter 2: Action, Practical Rationality and Problem Cases.................................................. 35  
Chapter 3: Arpaly's Theory of Practical Rationality................................................................. 66  
Chapter 4: Best Judgment and Practical Rationality................................................................. 102  
Chapter 4: The Kantian Theory of Practical Rationality, Revised......................................... 130  
References.................................................................................................................................... 163
Chapter 1: An Introduction to the Kantian Framework

Earlier, I suggested that the Kantian theory of action sees action as behavior that is genuinely guided by a series of representations which reflect the agent's understanding of what he is doing. I also suggested that the Kantian theory of practical rationality makes the claim that rational action is action that accords with one's best judgment, and irrational action as action that goes against one's best judgment - where acting in accordance with one's best judgment is taken to be a necessary condition of rational action, and acting against one's best judgment is taken to be a sufficient condition of irrational action. In this chapter, I will provide a more complete account of what Kantians commit themselves to when they make these claims. I will start out by presenting the Kantian theory of action, and then show how this theory of action serves as the basis for the aforementioned claim made by the Kantian theory of practical rationality. Now, in presenting the Kantian theory of action, there are a large number of Kantians that we can appeal to. That said, I think the best thing to do is to go to the source and start with Kant.¹ As I see it, there are a number of key passages found in the *Groundwork*, the *Critique of Practical Reason* and the *Critique of Pure Reason* that give

¹ Obviously, there is a sense in which we can't really go “straight to the source”, since understanding Kant is a complicated interpretive affair. There is also an intimidating amount of literature devoted to interpretation of Kant's works on ethics and practical rationality. That said, much of what I have to say in this chapter will be relatively uncontroversial. In places where I feel my claims are a bit off the beaten path, I will provide proper textual evidence for my conclusions.
us insight into how he views action. 2 I will show how he takes action to involve deliberation, best judgment and choice, and also explain how he takes consciousness to be involved in all three. The focus on consciousness may initially seem irrelevant, but it will become apparent in later chapters that the role of consciousness in action is a key point of contention.

After presenting Kant's theory of action, I will look at an enriched version of Kant's theory of action offered by Barbara Herman in *The Practice of Moral Judgment*. 3 As Herman takes herself to be giving a “normative reconstruction” of Kant, Herman's theory is a good example of a modern Kantian theory of action. The Kantian theory of action, as offered up by Kant and Herman, will serve as the basis for the claim made by the Kantian theory of practical rationality which I will present at the end of the chapter. The Kantian theory of action and the claim about rational and irrational action made by the Kantian theory of practical rationality, presented in this chapter, will be the subject of the rest of the dissertation.

**The Beginning of Kant's Theory of Action**

Kant's most clear discussion of action is contained in a somewhat esoteric passage regarding the actions of angels 4 and humans. Here he notes:

---


4 In other passages, Kant implies that he is talking about angels. However, he does not mean to suggest that angels exist – angels are considered a theoretical being that he uses for the purpose of illustrating how he views human action.
“If reason infallibly determines the will, the actions of such a being that are *cognized* as objectively necessary are also subjectively necessary, that is, the will is a capacity to choose only that which reason independently of inclination *cognizes* as practically necessary, that is, as good. However, if reason solely by itself does not adequately determine the will; if the will is exposed also to subjective conditions (certain incentives) that are not always in accordance with the objective ones; in a word if the will is not in itself completely in conformity with reason (as is actually the case with human beings), then actions that are cognized as objectively necessary are subjectively contingent, and the determination of such a will in conformity with objective laws is necessitation.”

Kant states here that the will of an angel is “a capacity to choose only that which reason independently of inclination cognizes as practically necessary, that is, as good.” Roughly, what he means here is that angels are such that they only choose that which they cognize *they ought to do*. According to Kant, angels only choose that which they cognize they ought to do because the choice of an angel is *immediate*, in the sense that “subjective conditions” like incentive do not affect the transition between cognition and choice. With regards to the above passage, one should also note that for Kant, angels have wills that are constructed such that choice is inherently *responsive* to their cognition of what they ought to do. Choice proceeds, for them, only *given a cognition of what they ought to do*. This, I think gives us a clue as to how Kant sees action. As I read Kant, action is behavior that results from *choice in response to cognition* of what one ought to do.

Human agents are different in that, given a cognition of what they ought to do, they don't necessarily choose that which they cognize they ought to do. The reason for this is evident. Human agents are exposed to subjective conditions that can cause them to choose otherwise. As I read Kant, it still is the case that human agents only
choose *given a cognition* of what they ought to do. We can see this is the case because, in his discussion of the aforementioned subjective conditions, he speaks of such conditions as playing an *accessory* role. Hence, when Kant states that subjective conditions come into play, he notes that the will is “exposed *also*” to subjective conditions, suggesting that the will is still exposed to the agent's cognition of what he ought to do. Given this, we can see that choice is still inherently responsive to the agent's cognition of what he ought to do. What can happen (and this is what makes us *human* agents) is that subjective conditions can affect the transition from the agent's cognition to his choice, and in such cases the choice of the agent can go awry. For both angels and humans, action is behavior that results from choice in response to cognition of what one ought to do. The difference is that subjective conditions can affect the choice of humans whereas it doesn't affect the choice of angels.

**Is Awareness of What One Ought to Do Required for Action?**

Keeping this in mind, it's evident that Kant takes action to necessarily involve cognition of what one ought to do. However, one might wonder whether *awareness* (by which I mean *conscious* awareness), of what one ought to do is required for action as well. One might argue that awareness of what one ought to do is required for action

---

6 Here, it's important to note that Kant takes cognition of what one ought to do as a requirement for action *in general*, not merely for *rational action*. This is evident, because he speaks of this cognition as being a condition for the action of human agents who act rationally *and* irrationally.

7 Henceforth, I will use the term “awareness” in place of “conscious awareness”. I do this partly to retain consistency with historical terminology, partly for the sake of convenience, and partly because I think the term “conscious awareness” might be redundant. There *is* good motivation to think that there is such a thing as “conscious awareness” as opposed to “unconscious awareness”. For instance, one might take the phenomenon of “blindsight” to be evidence of unconscious awareness, and hence one might consider the term “conscious awareness” to be non-redundant. That said, it's not clear to me that blindsight is an example of awareness without consciousness as opposed to a special case of responsiveness without awareness. (It's not even clear to me that it's a case of awareness without
because cognition implies awareness, and if cognition of what one ought to do is required for action, then awareness of what one ought to do must be required as well. However, this isn't entirely obvious, especially given that the term “cognition” is ambiguous. After all, Kant might see cognition as akin to belief, and believing that X doesn't necessarily imply awareness of X. I can believe that the sky is blue, and this doesn't necessarily imply that I am also aware that the sky is blue. Kant doesn't ever directly say that awareness of what one ought to do is required for action. However, there are some passages that support this claim. For instance, if we look at the *Critique of Pure Reason*, we find passages suggesting that *cognition implies awareness*. Here Kant states:

“There is no lack of terms suitable for each kind of representation... their serial arrangement is as follows. The genus is representation in general (repraesentatio). Subordinate to it stands representation with consciousness (perceptio). A perception which relates solely to the subject as the modification of its state is sensation (sensatio), an objective perception is knowledge (cognitio).”

This categorization of representations is interesting because Kant takes cognition (cognitio) to be a *type* of representation, in particular a subcategory of a type of representation which he calls “representation with consciousness” (perceptio). In light of this, one might see this as good reason to think that Kant takes cognition to imply awareness. If cognition does imply awareness, we get the result that cognition of what one ought to do inherently involves awareness of what one ought to do as well. Moreover, given that cognition of what one ought to do is required for action, we get the result that awareness of what one ought to do is required for action also.

In case one doesn't take this to be sufficient evidence for Kant's views on

---

8 CP, A:319.

consciousness, as opposed to a case of awareness without *visual* consciousness...), but I digress.
awareness of what one ought to do, we can point to a number of other passages. For instance, in the Lectures in Ethics, Vigilantius notes that Kant talks about free action as *that which can be imputed*. Vigilantius cites Kant as stating that “nothing can be imputed, save what is subject to laws, and in respect of which we are obligated *ad aliquid omittendum vel commitendum*. For only an action that rests on freedom is imputable, and freedom itself is nothing more than the capacity to be held accountable.” Here, when Kant speaks of imputation, he refers to both the ascription of action and the ascription of responsibility. Actions can only be imputed to an agent when he is *subject to laws that obligate him* to do or not do something – laws telling him how he ought to act. Later on, Kant makes it evident that being “subject to laws” involves awareness of the law that obligates, or awareness of what one ought to do. Specifically he notes that intentional immoral action is defined as action “undertaken after prior consideration with awareness of the law forbidding it.”

Kant's statements about imputation and awareness here might seem to apply specifically only to *immoral action*, but when one looks at the context of the passage, it

---

9 Roughly: obligated to someone, to do or not do.
11 LE 27:568. Kant's statements here are complicated. He contrasts an agent who performs an action “with intent” - that is, action “undertaken after prior consideration with awareness of the law forbidding it”, with unintentional action - presumably action undertaken with no awareness of the law forbidding it. The agent who acts unintentionally is specifically an agent who acts *while drunk* – an agent who is too drunk to be aware of the laws forbidding his actions. In the first case, he says that the actions can be imputed *directly*, whereas in the second case, he says that the actions can be imputed only *indirectly*. These statements might give the appearance that Kant does *not* in fact take “awareness of the law” to be a requirement for action, since he states that the unintentional actions (the actions without awareness) of the drunkard can be imputed *indirectly* to him. However, in an earlier passage (LE, 27:289) he clarifies that *technically speaking* the unintentional actions of the drunkard cannot be properly imputed to the drunkard. What *is* imputable to the drunkard is the action of *becoming drunk* – presumably an action that was undertaken *with awareness* of the law forbidding it. So, when he later says that the actions of a drunkard can be imputed to the drunkard *indirectly*, he means that the actions of the drunkard are essentially *signs of* another action that can be properly imputed to the agent. Kant still takes a very specific awareness to be a requirement for intentional immoral action.
becomes apparent that Kant's statements apply to action in general. Kant prefaces the passage with a statement that he is going to outline conditions that warrant lessening or withholding imputation of action – he does not specify immoral or morally good action. So, it is reasonable to conclude that Kant's statements about awareness are meant to apply to action across the board. Keeping this in mind, we have another instance in which Kant seems to speak of action as necessarily involving awareness of what one ought to do.

That said, we have now established one way in which Kant taking consciousness to be involved in action. Specifically, Kant takes it to be the case that awareness of what one ought to do is required for action.

**How to Get to Cognition of What One Ought to Do, Part 1**

I earlier argued that Kant takes action to require cognition of what one ought to do, but I haven't yet explained how Kant takes one to arrive at this cognition in the first place. In order to see how one comes to such a cognition, let us first look at his discussion of the application of the Categorical Imperative, contained in the first section of the *Groundwork*. Kant's primary goal in presenting this application is to show that the Categorical Imperative is a principle that has substantive results in determining right action. He wants to show that the Categorical Imperative is, in fact, the moral law. If the Categorical Imperative can yield the result that certain types of intuitively immoral action fail the Categorical Imperative test, then this is good evidence that the Categorical Imperative is the moral law. Kant's primary goal in these passages aside, I do think that there is a secondary reason for Kant's laying out such an application. As I see it, his
discussion also provides one of his first illustrations of how imperatives operate in 

human psychology. Here, let's look at Kant's discussion of the duty to not commit 
suicide:

“Someone feels sick of life because of a series of troubles that has 
grown to the point of despair, but is still so far in possession of his reason that he 
can ask himself whether it would not be contrary to duty to himself to take his 
own life. Now he inquires whether the maxim of his action could indeed become 
a universal law of nature. His maxim, however, is: from self-love I make it my 
principle to shorten my life when its longer duration threatens more troubles than 
it promises agreeableness. The only further question is whether this principle of 
self-love could become a universal law of nature. It is then seen at once that a 
nature whose law it would be to destroy life itself by means of the same feeling 
whose destination is to impel toward the furtherance of life would contradict itself 
and would therefore not subsist as nature; thus that maxim could not possibly be a 
law of nature and, accordingly, altogether opposes the supreme principle of all 
duty.”

According to Kant, maxims can be evaluated by subjecting them to the 
Categorical Imperative test. If such maxims fail the Categorical Imperative test, then one 
can conclude that it is morally impermissible to act on them. In this case, a maxim of 
suicide fails to pass the Categorical Imperative test. Since, intuitively, we take suicide to 
be morally impermissible, it appears that the Categorical Imperative test yields the 
correct result, in which case we have evidence that the Categorical Imperative is, in fact, 
the principle of morality.

Now, what I would like to point out is that Kant also hints that his illustration 
here is supposed to be an illustration of how he sees moral psychology proceeding. As I

---

12 G, 4:422.
13 For Kant, a maxim is a principle of action which specifies that one perform a type of action, under 
certain types of conditions, for the sake of certain types of ends. So, the maxim of suicide that Kant 
mentions in this passage is a maxim to end one's life, under the condition that further living promises 
more bad than good, for the sake of self-love. Maxims are to be thought of as general principles in the 
sense that they do not merely specify that one perform a specific token action, under a token 
circumstance, for a token end.
see it, if Kant's only purpose was to test the Categorical Imperative, then he would have illustrated the above deduction in the style of a math problem, or a hypothetical thought experiment in which he entreats the reader to imagine holding the maxim and imagine what it would be like if the maxim were a universal law. This is not, in fact, how Kant illustrates the case. Rather, he illustrates the deduction from the point of view of the agent. It is the agent who reflects on the maxim in question and then determines for himself whether his maxim is fit to be a universal law of nature (Kant notes that “he inquires whether the maxim of his action...”).

Given that Kant's illustration is meant to also be an illustration of the psychology involved in applying imperatives, we might think that Kant has outlined a three-step process. First, the agent is aware of a proposed maxim of action. Second, the agent determines for himself whether the maxim of action is one that satisfies the Categorical Imperative test. Third, after determining this, the agent then becomes aware of something new, he becomes aware that he ought (or ought not) act in accordance with the maxim in question.14 Of course, Kant's characterization here is a bit contrived. In particular, the second step may strike some readers as psychologically implausible, with virtue ethicists in particular questioning its psychological plausibility. In light of this, let me try to clarify Kant's views on each of these three steps.

Regarding skepticism of the second step, there are a few things to note. First, Kant's illustration here is specifically an illustration of only one version of the Categorical Imperative, the Formula of Universal Law. So, when he suggests that agents

---

14 Kant notes that after subjecting his maxim to the test, “it is then seen at once... that maxim... altogether opposes the supreme principle of all duty.”
determine for themselves if their maxim passes the Categorical Imperative test, he does not mean that agents explicitly invoke *this specific version* of the Categorical Imperative. Second, Kant later on makes it clear that the formulations of the Categorical Imperative are merely “ways of representing the principle of morality” and are simply meant to bring the “idea of reason closer to intuition (by a certain analogy) and thereby to feeling.”¹⁵ In light of this, one shouldn't think of *any* of the formulations of the Categorical Imperative as being literal word-for-word descriptions of the principle of morality that we invoke in practical reasoning. Rather, they should be read as his best attempts at making *more explicit* the principle of morality that we already invoke in practical reasoning.

Third, one must remember that Kant takes imperatives like the Categorical Imperative to be analogous to logical principles like modus ponens.¹⁶ Given this, it's reasonable to suggest that for Kant, the application of the imperatives works, from a psychological standpoint, in the same way as the application of modus ponens. We can make the principle of modus ponens explicit by laying out its form, and then consciously ask ourselves if a set of premises and conclusions conforms to it. However, this would be a caricature of how modus ponens operates in our psychology. We might rather suggest that we start out being aware of premises and a proposed conclusion, and through modus ponens, come to be aware that the conclusion is justified. I'm not particularly interested in the metaphysical status of the principle of modus ponens, but from a phenomenological-psychological standpoint this seems pretty accurate. Keeping this in

---

¹⁵ G, 4:436.
¹⁶ This is most evident in CPR where Kant lays out a table of categories that outlines the formal characteristics of practical judgment. Evidently, Kant sees practical judgment as being a kind of *logical* judgment, as this table of categories is analogous to the table of the *logical functions of judgment* given in CP and the table of the categories given in CPR.
mind, we might think that Kant takes the application of the Categorical Imperative (illustrated in the second step) to proceed in a similar manner. The agent starts out with awareness of a particular maxim and through the Categorical Imperative, he comes to be aware of how he ought act. Interpreting Kant's second step in this way, we can see that Kant's illustration of the application of the Categorical Imperative is psychologically plausible. He explains it in an overly explicit manner, in part, because he means to make the content of the Categorical Imperative more explicit. This doesn't mean that he takes agents to explicitly consult the Categorical Imperative in the evaluation of maxims.

Regarding the first and the third steps, one might be skeptical as to whether Kant really thinks that agents must be aware of their proposed maxim of action and aware that they ought or ought not act on it. It's important to make clear that Kant really thinks of the first and third steps in these ways because these steps make substantive claims about the awareness of the agent. That Kant thinks awareness is involved in these two steps becomes apparent when one looks at a few other passages found in the *Critique of Practical Reason* and the *Critique of Pure Reason*. In the *Critique of Practical Reason*, Kant asserts that the evaluation of maxims in accordance with imperatives is a function of the faculty of judgment.17 Moreover, in the *Critique of Pure Reason*, the faculty of judgment is illustrated as being a faculty whereby conscious representations are synthesized, by means of logical principles, into a unified conscious representation structured by the logical principles (I.e., a judgment).18 Keeping this in mind, we can see

---

17 For instance, in the *Critique of Practical Reason*, Kant lays out an illustration of the application of the Categorical Imperative, where he speaks of the application as involving the employment of a “rule of judgment”. In the same passage, he describes moral reasoning as a process of moral reasoning as being the exercise of “pure practical judgment”. CPR 5:68-69.

18 Kant's main discussion of this is found in the Transcendental Deduction of the *Critique of Pure Reason*, CP, B:139-148.
that when Kant talks about the evaluation of maxims, he does think that agents start out with an awareness of a maxim, and end with an awareness of whether they ought to act on the maxim or not. The evaluation of a maxim just is the exercise of the faculty of judgment whereby conscious representations (in this case, maxims) are synthesized by means of logical principles (the imperatives), into a unified conscious representation structured by logical principles (a cognition about whether one ought act on the maxim in question or not).

Keeping all this in mind, I think that we have a pretty clear picture of how Kant takes agents to evaluate maxims. An agent starts out with an awareness of a maxim, and through the faculty of judgment and imperatives, forms a judgment about whether he ought or ought not act on it, a judgment which carries with it awareness of whether he ought or ought not act on it. Of course, the initial project here was to explain how Kant sees agents coming to a cognition of what they ought to do. I have so far discussed how Kant takes maxims to be evaluated by imperatives. If the evaluation of maxims was the only process involved in coming to a cognition of what one ought to do, then my work would be done. However, as I read Kant, something more is needed. Kant seems to suggest that an additional process is involved in coming to a cognition of what one ought to do.

**How to Get to Cognition of What One Ought to Do, Part 2**

In order to see why coming to a cognition of what one ought to do involves more than the evaluation of maxims, we simply need to observe that agents oftentimes
arrive at a situation with the option of acting on more than one maxim. Given that the application of the Categorical Imperative to a single maxim only tells an agent whether or not he ought to act on that particular maxim, such an application isn't going to get an agent to a cognition of what he ought to do. After all, when Kant speaks of cognition of what one ought to do, he means cognition of what one ought to do all-things-considered. In the case that an agent has several maxims on hand, the application of imperatives to these disparate maxims cannot yield this all-things-considered cognition. Keeping this in mind, it seems that something else is involved in an agent's coming to a cognition of what he ought to do. What does Kant think is needed? We get a clue from some passages where Kant discusses the weighing of duties:

“But a wide duty is not to be taken as permission to make exceptions to the maxim of actions but only as permission to limit one maxim of duty by another (e.g., love of one's neighbor in general by love of one's parents), by which in fact the field for the practice of virtue is widened.”

Kant's main goal in this passage is to make clear the nature of wide duties. When Kant speaks here of “maxims of duty”, he refers to maxims that fail the Categorical Imperative test in a particular manner (they generate a contradiction in willing), thus yielding the conclusion that one ought to act on an opposing maxim. What's important to note here, however, is that Kant seems to imply that the application of the Categorical Imperative test does not yet yield cognition of what one ought to do. One has to make an additional judgment regarding how one ought to “limit one maxim of duty by another,” a judgment telling the agent which “maxim of duty” has priority in a

given situation. This reading of Kant is confirmed in the Metaphysics of Morals, where Kant discusses what he calls “competing grounds of obligation.” Kant notes:

“...a subject may have... two grounds of obligation, one or the other of which is not sufficient to put him under obligation, so that one of them is not a duty. When two such grounds conflict with each other, practical philosophy says, not that the stronger obligation takes precedence, but that the stronger ground of obligation prevails.”

Kant's main goal in this passage is to make sense of the concept of moral dilemma. Kant here means to reject the conclusion that moral dilemmas result from a conflict of obligations, since he thinks that a conflict of obligations is conceptually impossible. He instead suggests that moral dilemmas result from a conflict of grounds of obligation. This main goal aside, what's important to note here is that when Kant speaks of “grounds of obligation,” he essentially refers to “maxims of duty.” Keeping this in mind, Kant here directly says that the application of the Categorical Imperative test to a maxim may not put the agent under obligation – in other words it may not yield for the agent cognition of what he ought to do. In such a case, an additional judgment is needed, one which determines which of the two “grounds of obligation” prevails. Given this, we can see that Kant doesn't take the application of the Categorical Imperative to individual maxims to be the only thing involved in coming to a cognition of what one ought to do. As Kant sees it, the Categorical Imperative test might yield a ground of obligation when applied to an individual maxim, but ones does not yet cognize what one ought to do until an additional judgment weighing the grounds of obligation is made (determining which obligation is “the stronger ground of obligation”).

In light of this, we can see that Kant's practical psychology is a bit more

21 MM, 6:224.
complicated than one might realize. Given the quoted passages, Kant apparently sees two types of judgment being involved in coming to a cognition of what one ought to do. The agent may be aware of a maxim, and through the faculty of judgment and the imperatives, come to be aware of a *ground of obligation* telling him to act on that maxim. However after this, the agent then must make an additional judgment regarding how various grounds of obligation are to be weighed. Given that the faculty involved in both judgments is the *very same faculty of judgment*, we can assume that Kant takes the awareness of the agent in making this second judgment to be similar. Moreover, given that it is *imperatives* that supply the normative element of judgments, we can also assume that Kant takes imperatives to be involved in making this second judgment. Keeping this in mind, we can characterize this second judgment as follows. The agent starts out with an awareness of *grounds of obligation*, and through the faculty of judgment and the imperatives, arrives at a further judgment – a judgment about what he ought to do, all things considered. This second judgment gives the agent cognition of what he ought to do, all things considered, and an awareness of what he ought to do as well.22

**Kant and Choice**

So far, I have discussed Kant's take on cognition of what one ought to do. Of course, as I mentioned at the beginning of this chapter, Kant takes action to involve more than just cognition of what one ought to do. As I characterized it, Kant thinks of action

---

22 I should note at this point that Kant never explicitly outlines the deliberation process in this way. In light of this, we should think of what has been said as a reasonable reconstruction of how Kant might have viewed the entire deliberation process, given his statements in the *Groundwork*, *Lectures in Ethics*, and elsewhere. That said, there is excellent reason to think that the traditional characterization of Kantian moral psychology, as simply involving the direct application of imperatives to individual maxims, should be discarded.
as being behavior that results from *choice in response to* cognition of what one ought to do. In order to figure out how Kant thinks of choice, let's return back to the original passage that I cited at the beginning of the chapter.

“If reason infallibly determines the will, the actions of such a being that are *cognized* as objectively necessary are also subjectively necessary, that is, the will is a capacity to choose only that which reason independently of inclination *cognizes* as practically necessary, that is, as good. However, if reason solely by itself does not adequately determine the will; if the will is exposed also to subjective conditions (certain incentives) that are not always in accordance with the objective ones; in a word if the will is not in itself completely in conformity with reason (as is actually the case with human beings), then actions that are cognized as objectively necessary are subjectively contingent, and the determination of such a will in conformity with objective laws is necessitation.”

In this passage, Kant makes a few claims about choice. Angels have wills such that they choose only that which they cognize they ought to do. Humans are different in that they don't always choose that which they cognize they ought to do. Certain “subjective conditions” can come into play, causing them to choose something else. Moreover, as I have argued earlier, Kant here also suggests that choice is something that is inherently *responsive* to the agent's cognition of what he ought to do. Given this passage, it is still not entirely clear what choice itself is, and modern Kantians have had trouble making sense of it, from a metaphysical standpoint. Also, Kant at times uses the terms “choice” and “the will” interchangeably, and sometimes speaks of them as referring to the same thing, but from different perspectives. I am not so much interested in the metaphysics of choice or the will, so instead, let me try to say a bit about the psychology of choice – specifically how awareness is involved in choosing. One question that we might have in particular is whether agents are necessarily aware of the

---

23 G, 4:412.
24 MM, 6:213.
maxims that they choose to act on, when they choose them. In order to answer this question, let me turn to a passage, also contained in the *Groundwork*, where Kant discusses the nature of action. Here, Kant notes that:

> “Only a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles, or has a will”.

As we can see here, Kant provides another characterization of human action. Action, for humans, specifically consists in behavior that is determined in accordance with the representation of laws. This is to be contrasted with ordinary physical objects which act in accordance with laws, but not the representation of laws. Kant's discussion here is important, because he reveals in this passage a key fact about how he views choice. When agents choose, they do so in accordance with representations, specifically the representation of *laws or maxims of action*. Now, Kant's use of “representation” is significant because representations, as noted earlier in our discussion of cognition, have important connections to consciousness. Keeping this in mind, it would be helpful to determine if, in speaking of choice, Kant takes the representation of a maxim to necessarily involve awareness of the maxim as well.

Given a look back at Kant's categorization of representations, it may first appear that representation doesn't necessarily imply awareness. After all, Kant seems to suggest that there is a category of representations that are unaccompanied by consciousness. This is implied in Kant's categorization because the category of “representations with consciousness” (perceptio) is designated as a *subcategory* of

---

25 G, 4:412.
26 I should note here that Kant's use of the term “in accordance” is not to be confused with the use of “in accordance” when talking about best judgment, in the context of the traditional theory of practical rationality.
representations in general (repraesentatio). In light of this, one might be unclear as to whether the representation of a maxim implies awareness of the maxim, with regards to choice. However, it should be noted that in this section, Kant is specifically addressing a “happy-go-lucky confusion” that has resulted from a popular unscientific use of the term “idea.” As I see it, Kant attempts to clarify this confusion by substituting the term “representation” in for the term “idea” and then providing a comprehensive categorization of the different types of possible representations. Given that Kant's goal in this passage is to help clarify the way in which the term “idea” has been used by others, it's not clear that we should take Kant's discussion here of representation to be definitive. I think that a clearer view of Kant's position on this matter can be found in his discussion of pure apperception, located earlier in the Critique. Here Kant makes these remarks:

“The 'I think' must accompany all my representations, for otherwise something would be represented in me which could not be thought; in other words, the representation would either be impossible, or at least be, in relation to me, nothing...”27 (my emphasis)

Here, Kant states that the accompaniment of the “I think” of self-consciousness is necessary in order for a representation to be either possible or “in relation to the agent” something. Given that, for Kant, consciousness of an object really just is the object's being accompanied by self-consciousness, we can see that Kant does draw substantial connections between the concept of representation and the concept of consciousness. That said, one might think that this statement does little to help us when it comes to answering the earlier question. After all, it seems that even here, Kant is making room for the possibility that representations might exist without

27 CP, B:131.
consciousness, since they might exist but be “in relation to the agent, nothing”.

However, we should note that in the context of action, Kant's statement here actually does show that he takes agents to be necessarily aware of the maxims that they choose, when they choose them. As noted earlier, when Kant speaks of action, he specifically refers to behavior that can be imputed to the agent. Keeping this in mind, it might be the case that representations can exist without the accompaniment of the “I think” of self-consciousness. However, if a representation is “in relation to me, nothing”, then such a representation is irrelevant to an account of action. So far as a representation is relevant to action, Kant seems to suggest that awareness is required after all. Keeping this in mind, we have an answer to the question I posed earlier.

With regards to choice, Kant does take it to be the case that the representation of a maxim inherently involves awareness of the maxim as well. This is just to say that when an agent chooses to act on a maxim, he is aware of the maxim that he chooses to act on.

**A Modern Take on Kant's Theory of Action**

For the sake of clarity, let me review everything I have said about Kant's theory of action and the awareness of the agent in action. According to Kant, action is behavior that results from choice in response to cognition of what one ought to do. Cognition of what one ought to do necessarily involves awareness of what one ought to do. Coming to a cognition of what one ought to do involves the formation of two kinds of judgments. Formation of the first type of judgment involves awareness of proposed

maxims, and through the faculty of judgment and imperatives, coming to an awareness of grounds of obligation. Formation of the second type of judgment involves awareness of grounds of obligation, and through the faculty of judgment and imperatives, coming to an awareness of what one ought to do. Choice in response to one's cognition of what one ought to do involves awareness as well. Since one chooses in accordance with the representation of laws or maxims, and the representation of a maxim necessarily involves awareness of that maxim, choice necessarily involves awareness of the maxim that one chooses to act on.

Now, Kant's terminology is quite archaic, and adapting it to more modern terms will be helpful in later discussions. So let me give a more modern gloss on Kant's theory of action, one that I think will retain the substance of what has been said. Let me propose a few substitutions. First, when Kant speaks of an agent as being aware of proposed maxims of action, he means that the agent is aware of certain considerations as favoring acting in a certain way, where the considerations specifically invoke the end and circumstances given in the maxim. So, for instance, say that an agent is aware of a proposed maxim – I will eat, in the case that food is available, and I am hungry. To say that the agent is aware of this maxim is to say that the agent is aware of certain considerations (that food is available and that I am hungry) as favoring eating. I do not mean that the agent is aware that these considerations normatively support the action in question. Rather, I mean that the agent is simply aware of the considerations and the action as being material for normative practical reflection. Second, when Kant speaks

29 For a fuller explication of this idea, I suggest looking at Korsgaard's Locke Lectures where she discusses the concept of incentive.
of “grounds of obligation” he refers to what we now might call “reasons for action”. Here, when I say that an agent is aware of a reason for action, I mean that the agent is aware that certain considerations are reasons for action – that the considerations normatively support the action in question, in the context of a particular imperative. I do not merely mean that the agent is aware of something that happens to be a reason for action. Given these substitutions, we might have a less archaic characterization of the first type of judgment noted earlier. Formation of this first type of judgment involves awareness of considerations, and through the faculty of judgment and imperatives, coming to awareness of reasons for action.

As noted, Kant also suggests that an agent weighs grounds of obligation in coming to a cognition of what one ought to do. A cognition of what one ought to do is a essentially a judgment about what maxim one ought to act on, all things considered. Keeping this in mind, let me offer up another substitution. When Kant speaks of “cognition of what one ought to do”, he refers to what we now might call an “all-things-considered best judgment,” or simply a best judgment. Given this second substitution, we might have a less archaic characterization of the second type of judgment noted earlier. Formation of the second type of judgment involves awareness of reasons for action, and through the faculty of judgment and the imperatives, coming to an awareness of the reasons one ought to act on, all things considered. \(^3\) Again, this is not to be construed as merely being aware of reasons, where such reasons happen to be the ones that one ought to act on. The agent comes to awareness of a certain set of reasons as being specifically

---

\(^3\) For the sake of brevity and clarity, instead of saying “awareness of the reasons one ought act on, all things considered”, I will simply say “awareness of the reasons one ought act on”. This is to be distinguished from non-all-things-considered judgments which simply given the agent awareness that a certain consideration is a reason for action.
the reasons that he ought to act on.

A few more substitutions will be helpful. Kant suggests that when agents choose, they choose a particular maxim of action. Here, let me suggest that when Kant says this, he means that agents choose to *act for a certain set of reasons*. To say that an agent chooses a maxim “do X for the sake of end E, under conditions C” is to say that the agent chooses to X for reason R – where reason R invokes end E and conditions C in the context of a particular imperative. For the sake of clarity, we should be careful to distinguish between several ways of characterizing choice, when one interprets what it means for an agent to *choose to X for reasons R*. There is a way of interpreting choice as being a faculty for *picking amongst a set of options*. When I speak of an agent choosing to X for reasons R, I do not mean that he chooses in this sense, rather, I mean that the agent *decides on or commits himself* to acting for reasons R. For Kant, an agent's choice to act for a reason signifies his *commitment* to act in a certain way for certain types of reasons. Keeping all of this in mind, we can make the following translation with regards to the awareness of the agent in choosing. When an agent chooses to act for a reason, he is necessarily aware of the *reason that he commits himself to act on*. Finally, one last substitution. When Kant speaks of the process of coming to a best judgment, he refers to what we now might call *deliberation*.

The translation scheme I have provided isn't perfect, but it will help us to rephrase Kant's theory of action in terms that are much more easy to understand. As a review: Action is behavior that results from *choice in response to best judgment*. Coming to a best judgment involves the formation of two kinds of judgments. Formation of the
first type of judgment involves awareness of *considerations*, and through the faculty of judgment and imperatives, coming to an awareness of *reasons for action*. Formation of the second type of judgment (i.e., best judgment) involves awareness of *reasons for action*, and through the faculty of judgment and imperatives, coming to an awareness of *the reasons one ought to act on*. Choice in response to best judgment involves awareness as well. When one chooses, one chooses to act for a certain set of reasons, and for Kant this necessarily involves awareness of *the reasons one commits oneself to act on*. In all of these cases, awareness is to be taken in the de dicto sense, not the de re sense. So, for instance, when I say that the agent is aware of a reason for action, I mean that he is aware of the reason *as being a reason for action*. When the agent is aware of the reasons that he ought to act on, I mean that he is aware of such reasons *as being the reasons he ought to act on*. When I say that the agent is aware of the reason that he commits himself to act on, I mean that he is aware of the reason *as being the reason that he commits himself to act on*.

**A Modern Take – Herman's Theory of Action**

Now that we have a working version of Kant's theory of action, we can turn to a more modern version of Kant's theory of action, as presented by Barbara Herman. Before I get to Herman's account outright, I think it will be helpful to get some context. I take most of her account from her book *The Practice of Moral Judgment.*[^31] In this book, Herman focuses specifically on *deliberation* – the process whereby one comes to a best

[^31]: [PMJ](PMJ). The chapters of greatest significance are “The Practice of Moral Judgment” and “Agency, Attachment and Difference”
judgment. Moreover, much of what she says is limited specifically to what she calls *moral* deliberation. Her book is, in part, a response to a number of attacks launched by virtue ethicists regarding the psychological plausibility of Kant's *ethical* theory. Given this, some of my characterization of Herman's views will be a bit speculative, though there is good reason to think that her views on deliberation are not substantively different from her views on moral deliberation. That said, Herman is interested in what she calls a “normative reconstruction” of Kant, specifically an expansion and reinterpretation of Kant. Keeping this in mind, Herman's work provides us with a good modern version of Kantian action theory since her work is more constructive than revisionary. In what follows, I will try and keep clear on exactly how her views represent an expansion of Kant's theory of action.

**The Agent's Conception of the Good**

In order to get started, I think it will help to look at Herman's account of the agent's “conception of the good.”\(^{32}\) Now, Herman does not lay out a systematic account of the agent's conception of the good. Moreover, her account of this concept evolves over time, making a coherent account difficult to obtain. However, I think we can take the agent's conception of the good to consist of what we might call the agent's *value judgments* – specifically judgments about reasons for action and their comparative strength.\(^{33}\) Where do these value judgments come from? Well, Herman seems to suggest

---

32 PMJ, 193.
33 I apologize for the use of the term “value judgments” here, as there are many who would consider “value judgments” to include more than just the class of judgments about reasons for action. The term “value” in this context is borrowed from Herman, and so I use the term partly to maintain consistency with this particular body of Kantian literature.
that these value judgments are a kind of byproduct of the agent's deliberation in coming to best judgments on action. The idea is that when an agent deliberates, the judgments that he forms are *retained* after deliberation has finished, and these judgments form what she calls the agent's conception of the good.

To illustrate, say that a deliberating agent applies an imperative to the maxim “I will eat food when I am hungry,” on a particular occasion, and say that he comes to the judgment that he has a *reason to eat* – namely that hunger constitutes a reason to eat. In this case, the judgment that the agent forms about reasons for action will be one that is retained by him as part of his conception of the good. His conception of the good will contain a value judgment that says that hunger constitutes a reason for eating. For the sake of clarity, we should note that the value judgment that is formed as a result of this deliberation is *purely general*, in that it specifies as a general rule that hunger constitutes a reason for eating – it does not merely specify that *my hunger now* constitutes a reason for *my eating now*. Now, say that the agent has reasons that support acting in two different ways, and say that he comes to the best judgment that he ought to act on the first set of reasons instead of the second. In this case, the agent's conception of the good will contain a further *comparative* value judgment (one that is also general) – one that says that the first set of reasons has greater weight than the second. Since the value judgments that make up the agent's conception of the good are the product of the agent's deliberation, we can expect that a normal human agent will have a rich conception of the good in that it will be comprised of a great number of value judgments about reasons for action and their comparative weight.
What role does the agent's conception of the good play in Herman's account? Well, as I see it, the agent's conception of the good is designed to help solve a problem with Kant's theory of action. We can see what the problem is when we consider one way of looking at Kant's theory. As I noted earlier, Kant suggests that coming to a best judgment involves the agent's forming two types of judgments. In forming the first type of judgment, the agent starts out with an awareness of considerations and ends with an awareness of reasons for action. In forming the second type of judgment, the agent starts out with awareness of reasons for action and ends with awareness of the reasons he ought to act on. Given the kinds of awareness required by each of these types of judgments, one might think it psychologically implausible to suppose that agents actually form all of these judgments whenever they reach a best judgment on action. Sometimes agents simply arrive at a situation immediately aware of what they have reason to do. Keeping this problem in mind, Herman's “conception of the good” helps to bypass this problem by reducing the need for deliberation in coming to a best judgment on action. If the agent retains value judgments from prior instances of deliberation, then the entire process of deliberation will not always be necessary in reaching a best judgment. The requisite judgments will oftentimes already be contained in the agent's conception of the good.

The Role of the Agent's Conception of the Good

Of course, I haven't yet illustrated how the agent's conception of the good reduces the need for deliberation in coming to a best judgment. How does it play this role? Well, according to Herman, the agent's conception of the good interacts with two
other components – the agent's *perception of his situation* and the agent's *principles of salience*. When Herman speaks of the agent's perception of his situation, she refers to the agent's awareness of non-normative facts in his given situation. When Herman speaks of principles of salience, she refers to a special kind of knowledge that the agent has regarding when elements of his conception of the good are relevant to his situation.

Now, according to Herman, the agent's perception of his situation combines with his principles of salience and his conception of the good, *in an unconscious way*, to make the agent *immediately aware* of reasons for action and their comparative weight. So, for instance, an agent in a given situation might notice a variety of facts, among them that a friend is asking for money. The agent may have a principle of salience which states that when friends ask for money, this is a situation in which a judgment regarding how need constitutes a reason to act is relevant. In such a case, the principle of salience then combines with this noticed fact and the relevant value judgment to make the agent *immediately aware* that he has *a reason to act* – namely a reason to help his friend. Of course, the agent will likely have a complex conception of the good, a large body of principles of salience, and his perception of his situation will be complex. So, the agent may immediately come to be aware of a variety of reasons to act.

This example is one that illustrates how the agent's conception of the good can bring the agent to awareness of *reasons* for action. However, Herman also suggests that the agent's conception of the good can make the agent immediately aware of the *comparative strength* of reasons as well. To illustrate, an agent might have, in his conception of the good, value judgments stating that personal health constitutes a reason
to eat seafood, and that personal enjoyment constitutes a reason to eat steak. He might also have a comparative value judgment stating that personal health is more important than personal enjoyment. Given his perception of his situation and his principles of salience, the agent's conception of the good might lead the agent to become immediately aware that he has a reason to eat seafood, a reason to eat steak and that he has more reason to eat seafood than he has to eat steak. Keeping these illustrations in mind, we can see how the agent's conception of the good can lessen the role of deliberation in coming to a best judgment on action. Given that the agent's perception of his situation and his principles of salience can combine to make the agent immediately aware of reasons for action and their relative weight, it can oftentimes be the case that an agent does not need to engage in the entire process of deliberation in order to come to a best judgment.

There are a few observations that I should make here. Since an agent's conception of the good might not always have the relevant value judgments contained in it, the process of coming to a best judgment on action will not always be simple. So for instance, in the case that an agent doesn't already have a value judgment regarding how a consideration constitutes a reason for action, the agent's conception of the good will be of no use. In such a case, the agent will need to examine and evaluate considerations in order to come to a judgment about reasons for action. Likewise, in the case that an agent doesn't have a comparative value judgment in his conception of the good regarding a set of reasons for action, the agent's conception of the good will be of no use as well. In such a case, the agent will need to form a judgment on which of his reasons for action is
stronger. Of course, none of this introduces a significant problem for Herman's account. If anything, this makes Herman's account more psychologically plausible. Sometimes coming to a best judgment is more psychologically taxing than at other times, and Herman's account makes sense of why this is the case. The amount of “work” that the agent needs to do in forming a best judgment will depend on how expansive his conception of the good is.

We should also note that for Herman, although the agent's conception of the good can help a great deal in getting the agent to a best judgment on action, it does not get an agent immediately to a best judgment on action. At most, it can provide the agent with an awareness of reasons for action and an awareness of their comparative weight. In light of this, it is still required that an agent *form a best judgment on action*. This is to say that even for Herman, the exercise of the faculty of judgment can never entirely be eliminated in one's coming to a best judgment on action. As Herman sees it, formation of a best judgment *minimally* involves one's being aware of reasons for action, aware of their comparative weight, and given the faculty of judgment and the imperatives, coming to an awareness of the reasons one ought to act on. I point this out because it is apparent that awareness still plays a heavy role in Herman's theory of deliberation.

Incidentally, there is another sense in which awareness is implicated in Herman's theory of deliberation. Although the agent's conception of the good can reduce the need for forming judgments in coming to a best judgment, it still is the case that the value judgments contained in the agent's conception of the good are the product of *past deliberation*. So, for instance, an agent might already have a value judgment in his
conception of the good about a particular reason for action. However, the agent can only have this value judgment in his conception of the good in the case that he has previously formed a judgment about that reason for action at a prior time. Given this, it must have been the case in the past that the agent was aware of a consideration, and through judgment and the imperatives, came to be aware of a reason for action. So, there is a kind of “past awareness” requirement implicit in Herman's account of deliberation as well.

A Review of the Kantian Theory of Action

So far in this chapter, I have discussed Kant's theory of action and Herman's theory of deliberation. According to Kant, action is behavior that results from choice in response to best judgment. Coming to a best judgment requires the formation of two types of judgment. Formation of the first type of judgment involves awareness of considerations, and through the faculty of judgment and imperatives, coming to an awareness of reasons for action. Formation of the second type of judgment involves awareness of reasons for action, and through the faculty of judgment and imperatives, coming to an awareness of the reasons one ought to act on. Finally, choice involves awareness of the reason that one commits oneself to act on.

Herman supplements Kant's theory of action with her own account of the agent's conception of the good. Given that agents can retain value judgments in their conceptions of the good, agents will not always need to engage in the entire process of deliberation in coming to a best judgment on action. The agent's conception of the good
can make the agent immediately aware of reasons for action and their relative weight. However, Herman still thinks that in coming to a best judgment, agents minimally must still be aware of reasons for action and aware of their comparative weight. Moreover, elements of the agent's conception of the good are formed through prior deliberation, so there is a kind of “past awareness” requirement associated with action as well.

A Kantian Theory of Practical Rationality

Now that we have a Kantian theory of action in place, I can lay out the Kantian theory of practical rationality, and specifically its take on rational and irrational action. Getting to the Kantian take on rational and irrational action is a bit difficult. Kant doesn't actually seem to use the terms “rational action” and “irrational action” when discussing the evaluation of actions, and in fact such terms were not used in the literature of his times. That said, we can get a sense of how modern Kantians might construe rational and irrational action by looking at some of Kant's statements in the *Groundwork*. Here, we can get a theory that is based on Kant's own writings. Let me return to a previously quoted passage:

“If reason infallibly determines the will, the actions of such a being that are cognized as objectively necessary are also subjectively necessary, that is, the will is a capacity to choose only that which reason independently of inclination cognizes as practically necessary, that is, as good. However, if reason solely by itself does not adequately determine the will; if the will is exposed also to subjective conditions (certain incentives) that are not always in accordance with the objective ones; in a word if the will is not in itself completely in conformity with reason (as is actually the case with human beings), then actions that are cognized as objectively necessary are subjectively contingent, and the determination of such a will in conformity with objective laws is necessitation.”

34 G, 4:412.
As I noted earlier, when Kant talks about cognition of practical necessity, he essentially refers to best judgment. Given this, one thing that we can learn from this passage is that Kant speaks of best judgment as representing that which “reason... cognizes as practically necessary, that is, as good.” (my emphasis). As Kant sees it, best judgment represents reason's command on what one ought to do. Another thing to point out is that in this passage, Kant clearly distinguishes between two different types of agents based on how they choose. Angels are beings who choose “that which” they judge best, whereas humans, given the imposition of incentives, sometimes choose something other than “that which” they judge best.

Keeping these distinctions in mind, we can draw out a relatively simple theory of rational and irrational action: Rational action is action that results from one's choosing that which one judges best, and irrational action is action that results from one's choosing something other than that which one judges best. Since best judgment is primarily a judgment about the reasons one ought to act on, all things considered, we can further specify rational and irrational action in the following way: Rational action is action that results from one's choosing to act on the reasons picked out by one's best judgment, and irrational action is action that result from one's choosing to act on reasons that are not picked out by one's best judgment. Here, when I say that a reason is “picked out” by one's best judgment, I mean that the agent's best judgment designates that reason (or set of reasons) as the reasons he ought to act on. Moreover, there are multiple ways in which we can interpret this claim about rational and irrational action, but given the scope of the dissertation, let me interpret it in a very minimal way: It is a necessary condition of
rational action that an action be the result of one's choosing to act on the reasons picked out by one's best judgment, and it is a sufficient condition of irrational action that an action be the result of one's choosing to act on reasons that are not picked out by one's best judgment.

This characterization of the claim made by Kant's theory of practical rationality is interesting, because I think that it corresponds to a modern theory of practical rationality that has been endorsed and cited frequently in the literature, a theory of practical rationality that one might call the *traditional theory of practical rationality*. According to the traditional theory of practical rationality, rational action is action that accords with one's best judgment, and irrational action is action that goes against one's best judgment, where we again are to interpret this in the following way: one's acting in accordance with one's best judgment is to be taken as a necessary condition for rational action, and one's acting against one's best judgment is to be taken as a sufficient condition for irrational action. Looking back at my characterization of the claim made by Kant's theory of practical rationality, we can see that it looks a lot like the traditional theory of practical rationality. The traditional theory suggests that rational action is action that accords with one's best judgment, whereas irrational action is action that goes against one's best judgment. The Kantian theory of practical rationality supplements the traditional theory by giving an interpretation of what it is to act in accordance with or against one's best judgment. One acts in accordance with one's best judgment when one's action results from one's choosing to act on reasons picked out by one's best judgment. One acts against one's best judgment when one's action results from one's choosing to act
on reasons other than the ones picked out by one's best judgment. In light of this, we can see that the claim made by the Kantian theory of practical rationality (at least as interpreted so far) is in fact a version of the claim made by the traditional theory of practical rationality. For the purposes of the rest of this dissertation, let us interpret the Kantian theory of practical rationality as *endorsing* the traditional theory of practical rationality, while also *supplementing* it with this aforementioned interpretation of acting in accordance with and acting against one's best judgment. Obviously, this way of looking at Kant's theory of practical rationality is more than a bit anachronistic. However, for the purposes of this dissertation, this shouldn't be a problem.

**A Look Ahead**

From this point on, I will call Kant's theory of action, supplemented with Herman's theory of deliberation, “the Kantian theory of action”. I will also call the modern gloss on Kant's theory of practical rationality, that is, Kant's theory of practical rationality as interpreted through the lens of the traditional theory's claim about rational and irrational action, “the Kantian theory of practical rationality”. As noted in the introduction, the Kantian theory of action and the Kantian theory of practical rationality are the subject of significant criticism in a well-known paper by Nomy Arpaly. In the next chapter, I will discuss two of these criticisms. The first concerns the phenomenon of *instinctive action*. In responding to the first criticism, I will explain why I think it is not an effective criticism. The second criticism, a criticism that involves a character named Sam, is more daunting and will be the focus of the rest of the dissertation.
In the previous chapter, I presented Kant's theory of action and Kant's theory of practical rationality. I then presented a modern gloss on these theories, which resulted in a Kantian theory of action and a Kantian theory of practical rationality. As I noted in the introduction, these theories are the target of significant criticism by Nomy Arpaly. The most daunting criticism concerns a character named Sam – an agent who allegedly acts rationally even though he acts against his best judgment. This criticism is directed specifically at the Kantian theory of practical rationality, and specifically the aforementioned claim about rational and irrational action, but as we will see in later chapters, the criticism has implications for the Kantian theory of action as well. Now, before I get to this main criticism, I should note that Arpaly has a separate criticism that concerns the phenomenon of *instinctive action*. In light of this, I would like to address this criticism first, and then turn to Arpaly's more general criticism of the Kantian theory of practical rationality after that.

**The Problem of Instinctive Action**

In order to see what the problem is, it will help to remember that Kantian theory of action carries with it some interesting claims about the *awareness* of an agent who acts rationally or irrationally. According to the Kantian theory of action, an agent
who performs an action will be aware in the senses that I outlined in the previous chapter. This is to say that, minimally, the agent will be aware of reasons for action, aware of the reasons he ought to act on, and aware of the reasons that he commits himself to act on. However, these claims are problematic. As Arpaly rightly points out, sometimes agents act rationally or irrationally without being aware of any of these things. Cases of instinctive action are examples of this.

**What is Instinctive Action?**

When I speak of instinctive action, I refer to a class of actions that can be called rational or irrational, but also seem to be unaccompanied by *any* sort of awareness of reasons for action, awareness of the reasons one ought act on, or awareness of the reasons one commits oneself to act on. Instinctive actions are to be distinguished from what we might call *gut reflex behaviors* – for instance, pulling one's hand away from a hot stove. Gut reflex behaviors are unaccompanied by any of the aforementioned awarenesses, however as I see it, such behaviors are *not actions*, and so do not belong to the class of instinctive *actions*. Moreover, when I speak of instinctive action, I do not mean to imply that the action results from a *natural instinct* – for instance, the survival instinct or the instinct to reproduce. While some actions that result from natural instinct are instinctive actions, I do not take it as a requirement of instinctive actions that they arise from natural instinct. When I speak of action as “instinctive”, I simply mean that the aforementioned awarenesses do not accompany the action. Arpaly presents some examples of what I call instinctive action:
“It is not a provocative view that an accomplished tennis player, for example, does not have time to deliberate on all her moves during a fast-paced game. Not only that, but given the complex factors to which she responds, she is unlikely to be able to reconstruct all the reasons for action after the game. However, even after the ball is served, we can legitimately judge her moves as rational... or criticize her for irrationality.”

“Think, for example, of a case in which James enters a room and walks quickly out of it... he does not know why he leaves, and only retrospectively does he come to the conclusion that he left because, in a manner not deliberated upon, he registered that the people in the room were hostile to him. Such an action may seem to us to be wise...”

Looking at the first example more closely, the instinctive action in question would be one of the particular actions that the tennis player (let's call her “Sally”) executes during the game – for instance, Sally's action of *slicing the ball*. Here, we should be careful to distinguish Sally's action of slicing the ball (an *instinctive action*) from a distinct but related action, the action of *acting on instinct*. Sally might, for example, deliberate and then choose to *act on instinct* – where this is to be construed as Sally's intentionally *ceasing to consciously reflect on action* for an extended period of time. This acting on instinct is to be distinguished from her *instinctive actions*, that is, the actions that she performs while in an unreflective state. Both the instinctive actions and the action of acting on instinct *could* be potentially attributed to Sally, but the action that we are interested in is specifically Sally's action of slicing the ball, her instinctive action. Furthermore, we should be careful to separate out the *reasons* and *best judgment* that Sally may have with regards to her slicing the ball, from the reasons and best

---

judgment that Sally may have with regards to her acting on instinct. Sally evidently has a number of reasons for slicing the ball – for instance, a slice is a slow-moving ball that gives her time to recuperate, and Sally wants time to recuperate. However, these reasons are distinct from the reasons she may have for acting on instinct – for instance, she knows that she makes mistakes when she over-thinks during the game, and she wants to minimize error. In this case, we are concerned with the reasons and best judgment that Sally has with regards to her slicing the ball, not her acting on instinct.

Keeping these distinctions in mind, we can now see why such examples of instinctive action might pose a problem for the Kantian. According to Arpaly, Sally performs an action that can legitimately be called rational or irrational. If this is the case, then the Kantian is committed to the conclusion that Sally must have at least been aware of her reasons for action, aware of the reasons she ought to act on, and aware of the reasons that she commits herself to act on. Yet, Sally doesn't seem to be aware of any of these things at all. There is a sense in which she really just acts *without thinking*. Given the lack of awareness on Sally's part, it seems that the Kantian must suggest that Sally's instinctive action wasn't an action after all. So, it seems that the Kantian is unable to accommodate the claim that Sally performs an action that can be legitimately called rational or irrational when she instinctively slices the ball. In fact, the Kantian is unable to accommodate the existence of instinctive actions in general. In light of this, what is the Kantian to do about such cases?

36 Of course, a similar point can be made with regards to best judgment, but for the sake of eliminating redundancy, I will move on.

37 Whether they can legitimately called rational or irrational is a point that is up for debate. However, I think that popular intuition suggests that such actions can legitimately be called rational or irrational. Kant even seems to suggest that such actions can be called rational or irrational.

38 We should note here that Herman's notion of the agent's conception of the good cannot help the Kantian
The Instinct Solution

Regarding this problem, Arpaly suggests one solution, a solution that I will call “the instinct solution.” According to the instinct solution, in cases of instinctive action, the instinctive action was caused by a prior action – the act of acting on instinct. This prior action, in turn, was caused by a prior deliberation and choice\(^\text{39}\) to act on instinct. Here, the act of “acting on instinct” is to be construed as the action of refraining from deliberate reflection in determining how one acts, thereby allowing one to instinctively act.\(^\text{40}\) Since the instinctive action was ultimately caused by deliberation and choice (a deliberation and choice to act on instinct), the Kantian can suggest that, in cases of instinctive action, the agent was in fact aware of reasons for action, the reasons he ought to act on, and the reasons that he commits himself to act on. He was just aware of these things when he deliberated and chose to act on instinct at this prior time. Moreover, the agent's instinctive action can be considered rational or irrational, based on whether or not the agent acted in accordance with his best judgment or against it, in his acting on instinct.

Let me illustrate by looking at Sally’s instinctive action of hitting a slice, and let us suppose that in this case, Sally acted rationally in hitting her slice. The instinct in accounting for instinctive action, since Herman still requires that agents be aware of reasons for action, aware of the priority of such reasons, and aware of a best judgment in acting.

\(^{39}\) As noted in the previous chapter, “choice” is not to be referred to as a faculty for picking amongst a set of options, but rather a faculty for committing oneself to acting in a certain way, or deciding on acting in a certain way. To say that an agent chooses to act for a reason is to say that the agent commits himself to acting for a particular reason. I retain the use of the term “choice” in order to maintain consistency with terminology used by Kantians.

\(^{40}\) One might wonder whether "acting on instinct" can be truly described as an action proper, given that it doesn't correspond to a direct (and outwardly) physical behavior. Regarding this, let me simply note that the action of "acting on instinct" is an action that is described by many in the sports, art and music world as being an important part of being a good athlete or artist. Learning how to act on instinct is quite difficult - in fact, it takes a significant amount of practice to be able to do well. However, it is something that good athletes and artists are able to deliberately execute.
solution would have it that, on a prior occasion, Sally was aware of a certain set of reasons for action, and came to the best judgment that she ought to act on instinct, for a certain set of reasons φ. She then, on this prior occasion, chose to act on instinct for these reasons φ, and proceeded to act on instinct. As a result of her acting on instinct, Sally then instinctively hit a slice during the game. Since Sally's hitting a slice was ultimately caused by her deliberation and choice to act on instinct, this solution can claim that Sally was aware of reasons for action, aware of the reasons she ought to act on, and aware of the reasons she commits herself to act on. She was aware of these things when she deliberated and chose to act on instinct. Since, in choosing to act on instinct, she chose to act on the reasons that were picked out by her best judgment, this solution can also claim that Sally acted in accordance with her best judgment and hence acted rationally.

There are a variety of reasons why this solution will not work. First, as Arpaly notes, one might think it implausible to suppose that all of an agent's instinctive actions are the result of a deliberation and choice to act on instinct. The fact of the matter is that sometimes agents simply “fall into” acting on instinct without deliberating and choosing to do so. Given this, it seems impossible for the instinct solution to explain away all of an agent's instinctive actions. The proponent of this solution would either have to suggest that many of an agent's instinctive actions are not in fact actions after all, or suggest that all of an agent's instinctive actions are preceded by a prior deliberation and choice to act on instinct. Both suggestions seem implausible. Second, the instinct solution doesn't seem to capture the intuition that the rationality of the agent's acting on

---

41 Forgive the shorthand. When I say that agents deliberate and choose to act on instinct, I mean that they deliberate and form a best judgment that they ought to act on a maxim of instinct and then choose to act on this maxim.
instinct can differ from the rationality of the instinctive actions that result from her acting on instinct. After all, Sally could be perfectly rational in her decision to act on instinct, and nevertheless instinctively hit a slice for all of the wrong reasons.

Third, the instinct solution doesn't seem to credit agents with the correct action. If the solution proposes that the agent deliberates and chooses to act on instinct, and that she doesn't deliberate and choose to perform the instinctive action itself, then it seems that the agent should be credited specifically with the action of acting on instinct and not with the instinctive action itself. Perhaps if the agent knew that her acting on instinct would, in the future, result specifically in her instinctive action, then we might credit the agent with the instinctive action. But in the absence of the agent's knowledge of this fact, it's hard to see why we would credit her with the instinctive action as well.

Suppose, for instance, that Sally previously deliberated and formed a best judgment that she ought to act on instinct for reasons $\psi$. Suppose that she also chose in accordance with this best judgment by choosing to act on instinct for reasons $\psi$, and as a result instinctively hit a slice. The instinct solution, it seems, is forced to suggest that Sally performed the action of acting on instinct and not the action of hitting a slice. This is problematic because intuitively, we want to say that Sally should be credited with the action of hitting a slice. Without specifically saying that Sally deliberated and chose to hit a slice, how could such a solution really credit her with hitting a slice? The instinct solution, I think, tries to solve the problem on the cheap.

There is a similar problem that relates to this previous one. Intuitively, when Sally instinctively hit her slice, she did it for a certain set of reasons, reasons that she was
only later able to reconstruct, if at all – for instance, Sally hit a slice because the slice is a slow ball that gives her time to recuperate, and she needed to recuperate. However, the reasons that Sally has for acting on instinct are definitely not the same reasons that she has for hitting her slice. If Sally deliberates and chooses to act on instinct, her reasons for doing so likely have to do with the fact that she makes errors when she over-thinks during a game, and she wants to reduce errors in her tennis game. If this is right, then it's not clear that the instinct solution can credit Sally with hitting her slice for the reasons that she has for hitting her slice, and it seems that we want to credit Sally with hitting her slice for these reasons. The problem here is that the instinct solution seems to credit Sally with the wrong reasons in acting.

So, the instinct solution doesn't seem to be of much help. Is there a better solution? I think a closer look at the third problem can give us a clue as to how to proceed. Essentially, the instinct solution attempted to get around the original problem by showing how the agent's instinctive action resulted from a prior instance of deliberation and choice – an instance where the agent deliberated and chose to do something else. The instinct solution suggested that the agent previously deliberated and chose to act on instinct. The third problem resulted because the action that the agent deliberated on and chose was different from the action that the agent was credited with. So long as there is a difference between the action that the agent deliberates on and chooses, and the action that the agent is credited with, we run into the third problem. If we try and save the day by suggesting that Sally previously deliberated and chose to build her character, and suggest that Sally's slice resulted from her good character, we
end up only attributing to Sally the action of *building her character*, not the action of hitting a slice. If we try and save the day by suggesting that Sally previously deliberated and chose to *shoot them all and let God sort it out*, and that Sally's slice resulted from this, we end up attributing to Sally the action of *shooting them all and letting God sort it out*, not the action of hitting a slice.

**The Maxim Solution**

This analysis of the problem does make a solution apparent. If the agent instinctively Xs, and the problem occurs because we say that agent previously deliberated and chose to do something *other than X-ing*, perhaps the solution then is that we ought to say that the agent previously deliberated and chose to *X*. Specifically, if the problem occurs because we say that Sally previously deliberated and chose to do something *other than* hitting a slice, perhaps the solution is that we ought to say that Sally previously deliberated and chose to hit a slice. Let me be a little more clear. I do not mean to suggest that Sally previously deliberated and chose to, in the future, hit a slice. What I mean to suggest is that on a previous occasion, Sally deliberated and chose to hit a slice (an action that she then executed), and that Sally's future slice resulted from this prior deliberation and choice as well.

What I am essentially suggesting is that multiple token actions can all result from a single deliberation and choice on the part of the agent. Admittedly, this solution might not seem to make any sense. For if an agent previously deliberated and chose to X in the past, then the result of this prior deliberation and choice would seem to be the
agent's X-ing in the past. Given this, it's hard to see how the agent's X-ing again in the future could result from this prior deliberation and choice. However, I think that resistance to this solution is based on a certain conception of choice, and the relation between choice and action. Under one sort of view, the object of the agent's choice is a specific token action. So, when an agent chooses to X, what he chooses is the specific token action of X-ing that he then performs. Here, the agent's action is attributed to the agent because the token action X is caused by the agent's choice, and the token action that we attribute to the agent is the very action that the agent chose. So, say that Sally chooses to eat a hamburger, and then proceeds to eat a hamburger. We say that Sally ate the hamburger because Sally's choice caused the token action of eating the hamburger, and the token action of her eating the hamburger is the token action she chose to perform. If we adopt this characterization of choice and the connection between choice and action, we can see why the aforementioned solution makes no sense. For if Sally performed a similar action in the future, this would be a different token action, and this later token action would not be the one that she chose in the past. Under this picture, multiple token actions could never result from the same choice.

This characterization of choice and the relation between choice and action is interesting, but we don't have to adopt it. If you're a Kantian, you think that what an agent chooses isn't a token action but rather a maxim of action, or a principle which describes both the action to be performed and the reasons for performing it – a principle of the form “perform action X, for reasons Y.” In the first chapter, I suggested that choosing a maxim amounts to choosing to act for a reason. This translation wasn't
entirely inaccurate, but it does mask a substantive claim that the Kantian makes about the object of choice. When one chooses, one doesn't simply choose a token action for a token reason. Rather, one chooses a principle of action which specifies that one *act in a certain way when a certain type of reason surfaces*. So, when an agent chooses to eat, he does not merely choose the token action of eating, he chooses a principle of action which specifies that *he eat when he wants to satiate his hunger*. Why is this important? Well, I think it's important because having a different conception of the object of choice makes it possible to construct a different picture of the relation between choice and action.

What might such a picture look like? Well, we might construct the connection between action and choice as follows: An action is attributed to an agent when the action is a manifestation of a disposition to act (a disposition of the form: perform action X, in the case that Y is present), the *disposition* itself results from the agent's choosing a maxim (a maxim of the form: perform action X*, for reasons Y*), and the disposition that is manifested in the agent's action “matches” the agent's chosen maxim.42 Under this view of choice and the connection between choice and action, it's easy to see how multiple token actions could all result from a single deliberation and choice on the part of the agent. A single deliberation and choice on the part of the agent can establish a disposition to act in the agent, and dispositions themselves can be manifested in multiple token actions.43

---

42 The disposition would “match” the maxim when $X = X^*$, and $Y = Y^*$.
43 Incidentally, this picture does seem to suggest an odd psychological account regarding how actions *in general* are produced. Specifically, one might think that in cases of *ordinary action* where the agent directly chooses to act, I am suggesting that the agent's choice results in *a disposition to act*, where this disposition *then manifests itself in the agent's action*. This picture might seem odd because it places one too many psychological components in between the agent's choice and his action. I actually don't find it *that* odd, but in the case that one finds it too odd, let me offer up a response. The Kantian can here suggest that choices cause *both* actions and dispositions to act. The account that I give above then is
Keeping all this in mind, how is it that we credit Sally with instinctively hitting her slice? Well, what I am suggesting is that on a prior occasion, Sally deliberated and came to the best judgment that she ought to hit a slice for a certain set of reasons \( \varphi \). She then chose in accordance with her best judgment by choosing to hit a slice for these reasons \( \varphi \). Sally's choice resulted in a disposition to hit a slice in the case that \( \varphi \) is present. Her disposition was manifested in her hitting a slice in the past, and her disposition was also manifested in her hitting a slice, instinctively, in the future.

Note that since Sally deliberated and chose to hit a slice, we can credit Sally specifically with the action of hitting a slice, not some other action. Moreover, this solution (henceforth the “maxim solution”) bypasses the other three problems associated with the instinct solution. Unlike the instinct solution, the maxim solution lays out plausible precursors to the agent's instinctive action. For Sally's slice is the result of a very refined set of dispositions, and it's hard to see how she could have gotten this great set of dispositions without previously working at it – a work that is plausibly the result of prior deliberation and choice. After all, a slice (or any other tennis shot) is a rather unnatural movement, one that takes a great deal of training to learn how to do correctly, and training requires a great deal of deliberation. Training is, after all, a complicated affair that involves a deliberative trial-and-error process. Of course, there is a sense in which an agent can be trained unconsciously to perform a limited number of actions (a la Pavlov), but it's plausible to suggest that such training doesn't yield action proper so much as reflex behavior. When it comes to instinctive action that is done for a reason, it's hard to see how an agent could have acquired her unique set of dispositions without a specifically an account of action in cases where the agent's action is the manifestation of a disposition.
specifically deliberative conscious training.

Unlike the instinct solution, the maxim solution also does not require that the rationality of the agent's instinctive action piggyback on the rationality of the agent's acting on instinct. Given that the instinctive action does not need to be caused by a prior deliberation and choice to act on instinct, the instinctive action can be judged on its own. Under this solution, an action can be instinctive but rational when it is a manifestation of a disposition that results from a choice in which the agent chooses to act on the reasons picked out by her best judgment. Similarly, an action can be instinctive but irrational when it is the manifestation of a disposition that results from a choice in which the agent chooses to act on reasons that are not picked out by her best judgment. Last, we can credit Sally with hitting her slice because of the reasons that she has for hitting her slice, not for the reasons that she has for acting on instinct. After all, when Sally previously engaged in deliberation and choice, she chose to hit her slice for a specific set of reasons (in other words, φ), and these are plausibly the very same reasons that she acted on when she instinctively hit a slice in the future.

Returning back to the problem of instinctive action, we can also see that the maxim solution provides us with a solution to this problem as well. Instinctive actions do not need to be immediately accompanied by awareness of reasons for action, awareness of the reasons one ought to act on, and awareness of the reasons one commits oneself to act on. This is because the deliberation and choice that ultimately causes the agent's instinctive action occurs at a significantly prior time. Moreover the agent's choice results in a disposition to act, and dispositions can operate independently of the agent's
deliberation and choice. In light of this, an agent like Sally can instinctively act without immediate awareness of reasons for action, awareness of the reasons she ought to act on, and awareness of the reasons she commits herself to act on. Given that one adopts the maxim solution, the Kantian can provide a plausible response to the problem of instinctive action.44

The Main Problem - The Case of Sam

At this point, we can now turn to the main argument that Arpaly presents against the Kantian theory of practical rationality. As noted, the Kantian theory of practical rationality states that rational action is action that accords with one's best judgment, and irrational action is action that goes against one's best judgment. What is Arpaly's argument? Well, Arpaly's argument consists of a putative counter-example. The counter-example involves a character named Sam, an agent who allegedly acts rationally even though he acts against his best judgment.

44 I should note that there is a problem that lurks for the solution I have provided to the problem of instinctive action. I don't address it in the main part of the chapter because I don't want to sidetrack the main discussion. Some might suggest that the solution I provide is a solution that explains only a certain subclass of instinctive actions – namely habitual actions. In light of this, they might contend that sometimes agents can act instinctively without having previously performed a similar action – their instinctive action represents the first time they have performed an action of that sort. Regarding this, let me make a few observations. First, in my discussions with philosophers and non-philosophers, it seems that the jury is still out on whether these non-habitual instinctive actions are genuinely actions. Intuitions are not decisively united one way or the other, and so it's not clear that these types of actions represent genuine counterexamples. Second, let me add that it's not clear to me that people genuinely have meaningful intuitions regarding cases like this in the first place. Part of this has to do with the fact that people are terrible at determining if, in fact, their actions or the actions of others are performed for the first time. In light of this, to claim that the general population has substantive intuitions about this specific subclass of instinctive actions strikes me as overconfident, and given this, I do not feel enormously inclined to respond to claims that intuition speaks in favor of including non-habitual instinctive actions as being genuine actions. In light of these observations, let me suggest that the jury is still out on whether these sorts of examples provide actual counter-examples to the solution I offer up. At the very least, I have provided a solution for the Kantian to address the phenomenon of habitual actions, and in light of the fact that the status of non-habitual instinctive actions is questionable, this is, for all intents and purposes a solution to the problem of instinctive action, generally speaking.
Who is Sam? Well, according to Arpaly, Sam is a college student who has previously tried to pass his exams by isolating himself from friends and studying feverishly. Unfortunately, this course of action has never lead Sam to be particularly productive. Instead of being productive, he gets lonely and instead plays video games. The result is that Sam never does very well on his exams when he isolates himself from the outside world. Sam comes to believe that his isolation never gets him the results that he wants, but he never reflects on this fact in any normatively significant way – he never notices how this fact might present him with a reason against isolation.

Before the end of the next semester, Sam notices that his exams are coming up sooner than expected, and this makes him anxious. He is aware of a variety of reasons for action, but he fails to remember that extreme isolation causes him to do badly on his exams. As a result, when he deliberates about what he ought to do, he again comes to the conclusion that he ought to study in isolation for a certain set of reasons $\phi$. However, Sam does not in fact proceed to study in isolation. Instead, Sam finds himself studying with his friends – a tactic that does not make him so isolated, but nevertheless makes him significantly more productive. Moreover, Arpaly states that unbeknownst to him, Sam acts for the best reasons, among them the reasons that he failed to attend to in his deliberation (let's call these “reasons $\psi$”). In fact, according to Arpaly, Sam acts on reasons $\psi$ because they are the best reasons that he has for action.\(^{45}\)

Now, according to Arpaly, Sam acts rationally. However, it seems that the Kantian theory of practical rationality must suggest that Sam acts irrationally. After all,

\(^{45}\) Arpaly actually does not use this exact terminology in describing the case of Sam. However, the terminology does line up with the terms Arpaly does use. Moreover, Arpaly has explicitly endorsed this specific characterization of Sam in conversation.
Sam deliberated and came to the best judgment that he ought to study in isolation for a certain set of reasons $\varphi$. He then ended up studying with friends for reasons $\psi$, an action that presumably resulted from his choosing to study with friends for reasons $\psi$.\footnote{As noted earlier, we should be careful to think of choice as simply being a faculty whereby one commits oneself to or decides on acting for a reason. This characterization of choice is to be distinguished from a much more involved characterization of choice which sees choice as inherently being a faculty for picking amongst a set of options. I am retaining the use of the term “choice” in part because of its use in the Kantian literature. Now, there is a potential problem with making the above assumption that Sam chooses to act for reasons $\psi$, as Kantians suggest that choosing to act for a reason requires awareness of the reason one commits oneself to act on. This is a problem that I will deal with in the final chapter.} If this is right, then Sam's action resulted from his choosing to act on reasons that are not picked out by his best judgment. Given this, the Kantian must suggest that Sam acts against his best judgment and hence acts irrationally. If we are to take Arpaly's case seriously, and I think we should, it seems that we must reject the Kantian theory of practical rationality. One can act rationally even when one acts against one's best judgment.

**Getting to the Problem**

What are we to make of this example? Well, there are actually two sorts of problems that we have here. The first problem, noted previously, is that the Kantian seems unavoidably lead to the conclusion that Sam acts irrationally. Given that Sam has already formed a best judgment that he ought to study in isolation for reasons $\varphi$, and given that his action presumably results from a choice to study with friends for reasons $\psi$, the Kantian must conclude that Sam acts against his best judgment and hence acts irrationally. The second problem is that the Kantian seems unable to account for how Sam could act rationally. In order to claim that Sam acts rationally, the Kantian must claim that Sam acts in accordance with his best judgment – that Sam's action results from
his choosing to act on the reasons picked out by his best judgment. However it seems that the Kantian cannot claim this at all. In order to claim this, the Kantian must at least claim that Sam has a best judgment that tells him he ought to study with friends for reasons $\psi$. Unfortunately, Sam doesn't seem to have such a best judgment on hand – Sam certainly isn't aware that he ought to act on reasons $\psi$. Moreover, Sam already has a best judgment that he ought to study in isolation for reasons $\phi$. In light of this, it seems that the Kantian cannot claim that Sam acts rationally.

We should also note that the maxim solution, proposed earlier, is not of much help. The maxim solution can assert that Sam previously deliberated and came to the best judgment that he ought to study with friends for reasons $\psi$, and then chose to act in accordance with his best judgment by choosing to act on reasons $\psi$. This choice caused a disposition in Sam to study with friends when $\psi$ was present, and this disposition was manifested in his eventual action. This would be a potentially promising explanation of Sam's action. However, there are problems with this story. First, it still is the case that Sam has a best judgment stating that he ought to study in isolation for reasons $\phi$. Given this, it appears that the proponent of the maxim solution is still forced to say that Sam acts against a best judgment, since Sam's action results from a choice to act on reasons that are not picked out by one of his best judgments (his current one). In light of this, the proponent of the maxim solution is forced to say that Sam is at least partly irrational in acting. Second, as Arpaly tells the story, Sam has never deliberated in coming to the best judgment that he ought to study with friends for reasons $\psi$, in which case the maxim solution is of no help at all.
Some Simple (and Unhelpful) Responses

Keeping all this in mind, here are two simple responses that one might propose. First, one might simply claim that Sam's action is irrational. This would, at first glance, appear to solve the problem for the Kantian quite quickly. Arpaly's case would not be a counter-example to the Kantian theory of practical rationality since the Kantian theory of practical rationality would correctly identify Sam as an agent who acts irrationally. Second, one might simply deny that Sam acts at all. Under this response, Sam's behavior would be deemed a mere gut reflex instead of an action. This would also, at first glance, appear solve the problem for the Kantian quite quickly. Arpaly's case would not be a counter-example to the Kantian theory of practical rationality since Sam doesn't count as an agent who acts rationally and also acts against his best judgment.

While these responses do have some initial appeal, I think that they put the Kantian theory of practical rationality in troubling territory. Why? First, one must note that many people have been convinced by the case of Sam. They consider Sam to be an agent who acts and moreover acts rationally. To deny these claims is to suggest that a rather large number of people have wildly incorrect intuitions about action and rational action. Given that one of the goals of philosophy of action is to capture our intuitive judgments about the actions of agents, it is difficult to accept this solution. Second, both solutions I think dismiss too early the possibility that the Kantian theory of practical rationality can be reconstructed in ways that both accommodate Arpaly's case and are acceptable to Kantians. Moreover these simple solutions dismiss the possibility that
*Arpaly has misinterpreted the Kantian theory of practical rationality* in describing the case of Sam. It seems presumptuous to deny these possibilities in advance of seeing how the Kantian theory might be reconstructed, especially when such attempts might be genuinely illuminating. Here, one should keep in mind that the work of many modern Kantians is specifically the result of a reconstruction of Kant's views in response to problem cases presented by virtue ethicists. In retrospect, their work proved to be worthwhile.

**The First Half of a Solution**

In light of what has been said, I think we should see what other options the Kantian has when it comes to accommodating the case of Sam. In the rest of this chapter, I will present a solution to the *first* of the two problems that I mentioned earlier. The solution takes a closer look at the concept of acting against one's best judgment, and more specifically the Kantian interpretation of this concept that I presented earlier. The solution will show that the concept of acting against one's best judgment is a bit more complicated than one might realize, and will also show that there are good reasons for Kantians to provide a more sophisticated interpretation of this concept. Moreover, the solution will show that Sam does not actually act against his best judgment after all. Now, this solution will show that the first problem isn't a problem for the Kantian theory of practical rationality. However, I think it will also show that one shouldn't reach so quickly to either of the simple solutions given earlier. A motivated revision of the Kantian theory of practical rationality *can* achieve substantive results in dealing with the
case of Sam. This should give us a kind of legitimate hope that the Kantian can successfully accommodate the case of Sam in the end.

As noted, the first problem concerns the fact that the Kantian theory of practical rationality seems forced to say that Sam acts irrationally. After all, Sam is an agent who seemingly acts against his best judgment, since his action presumably results from a choice to act on reasons that are not picked out by his best judgment. But does Sam really act against his best judgment? I think there are good reasons to think otherwise. To see why, let me discuss the concept of acting against one's best judgment in more detail.

**What it is to Act Against One's Best Judgment?**

What is it to act against one's best judgment? Well, at the end of the last chapter, I outlined how a Kantian might interpret this. One acts *against* one's best judgment when one's action results from one's choosing to act on reasons that are not picked out by one's best judgment. While I think this is a relatively intuitive account of what it is to act against one's best judgment, I think that it conflates two concepts – *acting against* one's best judgment, and merely *not acting in accordance with* one's best judgment. What is the difference between these two concepts?

Let me start with one way of making the distinction between these concepts. In particular, let us focus on an agent named Pam. Pam has previously deliberated and formed the best judgment that she ought to A for reason $\phi$. She also has, on this occasion, chosen to A for reason $\phi$ as well, and this choice has instilled in her an
unconscious disposition to A in the case that φ is present. Keeping this in mind, at a point in the future, when φ is present, she becomes aware of a variety of reasons for action (ψ, ω, λ, and δ) but fails to attend to reason φ when she deliberates. When I say that she fails to attend to reason φ, I simply mean that she fails to be aware of φ as being a reason for action. She may, as a matter of fact, be aware of φ, but because she fails to be aware of it as a reason for action, she subsequently does not take it up into her deliberation at the time. Given this, when she deliberates about what she ought to do, she comes to the best judgment that she ought to B for reasons ψ. However, before she chooses to B for reasons ψ, her unconscious disposition kicks in and she instinctively As for reason φ. Now, looking at Pam's action, it is evident that Pam's action is ultimately the result of her choosing to act on reasons that are not picked out by her current best judgment. After all, the choice that caused her action was a choice to A for reason φ. Her current best judgment is a best judgment to B for reasons ψ. Nevertheless, I think it is also the case that Pam has not acted against her current best judgment. Rather, she has merely not acted in accordance with her current best judgment.

In case one doesn't find this intuitive, consider another agent, Stan. Stan, unlike Pam, doesn't have the same unconscious disposition ingrained in him. He is aware of the same reasons as Pam, and comes to the same best judgment as well (that he ought to B for reasons ψ). However, unlike Pam, he directly and consciously chooses to act on one of the reasons that he rejects in his deliberation (say he chooses to C for reasons λ).

---

47 One might think that, since her disposition kicks in, (resulting in her A-ing for reasons φ) reasons φ must have registered with Pam in some significant way. I deny that this is the case, as dispositions can apparently operate independently of the consciousness of the agent.

48 We could also simply say that Pam acts on reasons φ, but I focus on the fact that Pam's action stems from her choosing to act on reasons φ in the past, because this will have relevance to a specific Kantian point about what it is to act against one's best judgment, later on in this chapter.
Now, as I see it, Stan is an agent who definitely acts *against* his current best judgment. Pam, on the other hand, seems like someone who merely does not act in accordance with her current best judgment. Keeping this in mind, if one takes both Pam and Stan to be acting against their best judgments, one does not have the resources to adequately explain what the significant difference between them is. The distinction between *acting against* and *not acting in accordance with* seems to be a good way of cashing out this difference.

Of course, the terms “acting against” and “not acting in accordance with” haven't been spelled out completely. Let me give a more substantive analysis, an analysis that I think will help my case. One might think that best judgments are formed in a particular context. Specifically, when agents form best judgments, they form them *given a set of reasons for action, \( \{\varepsilon, \zeta, \eta, \theta\} \). Given that best judgments are formed in a particular context, we might also think that best judgments have a *domain of authority* in the sense that best judgments only tell an agent what reasons he ought to act on, *given a particular domain of reasons for action, \( \{\varepsilon, \zeta, \eta, \theta\} \). I don't mean to suggest that the domain is part of the content of a best judgment. However, I do think that best judgments have a domain in the sense that they are relativized to a specific set of reasons for action. Keeping this in mind, we might think that we can further specify what it means to act against a best judgment and what it means to simply not act in accordance with it. One acts against a best judgment when one's action is the result of a choice to act on a reason that is not picked out by the best judgment, and the reason is *contained in the domain of* that best judgment. One merely does not act in accordance with a best judgment when one's action is the result of a choice to act on a reason that is not picked out by the best
judgment, and the reason is not contained in the domain of that best judgment.

How does this analysis help to legitimize the distinction I made? Well, one might think it plausible that acting against a best judgment involves acting against the authority of that best judgment. Keeping this in mind, the above analysis of the two concepts makes sense. When an agent's action results from his choosing to act on a reason not picked out by his best judgment and the reason is contained in the domain of his best judgment, there is a sense in which the agent has acted against the authority of his best judgment. This is because the reason the agent chooses to act on is contained in the domain of the best judgment, and so the best judgment has the authority to say that the reason in question is not the best one available. Here, we can say that the agent has acted against his best judgment. Alternatively, when an agent acts on a reason not picked out by his best judgment and the reason isn't contained in the domain of his best judgment, there is a sense in which the agent has not acted against his best judgment. This is because the reason that the agent chooses to act on is not contained in the domain of his best judgment, and so the best judgment has no authority to say that the reason in question is not the best one available. Here, it is legitimate to simply say that the agent has merely not acted in accordance with his best judgment.

Returning back to Pam and Stan, we can see why Stan has acted against his best judgment, whereas Pam has simply not acted in accordance with hers. Looking at the choice that causes their actions, Stan chooses to act on a reason that is not picked out by his best judgment, but he also chooses to act on a reason that is contained in the domain of his best judgment. This is evident because Stan chooses to act on a reason that
he is aware of in his deliberation. His best judgment has authority over it and has picked another reason as being better. In light of this, it is evident that Stan acts against his best judgment. Pam likewise chooses to act on a reason that is not picked out by her best judgment, but she chooses to act on a reason that is not contained in the domain of her best judgment. This is evident because her deliberation failed to attend to the reason that she chooses to act on. Her best judgment has no authority over it, and so it has no authority over whether she ought to act that reason or not. Given this, Pam does not act against her best judgment, she simply does not act in accordance with it.

A Further Kantian Motivation for the Distinction

This account of what it is to act against one's best judgment does help to make sense of Pam and Stan. However, I think that Kant provides an additional characteristic that distinguishes acting against a best judgment from merely not acting in accordance with it. In order to see what this characteristic is, I suggest that we go back to a passage that I quoted in the previous chapter:

“Everything works in accordance with laws. Only a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles, or has a will. Since reason is required for the derivation of actions from laws, the will is nothing other than practical reason.” (my emphasis)\(^49\)

Kant's statement here is partially designed to reveal one aspect of the will – the will is the faculty of judgment/reason, in its practical employment (i.e., the will is practical reason). However, the quote here also shows how Kant sees the relationship between choice and best judgment. In this passage, Kant states that actions are derived

\(^{49}\) G, 4:412.
from laws by means of the faculty of reason (essentially, the faculty of judgment). He then follows this statement with a description of the role that the faculty of reason plays in the transition between best judgment and choice. In light of this, Kant appears to take the choices and best judgments of an agent to have a special psychological relationship with one another – choices are derived from best judgments by means of the faculty of judgment, much in the same way that an agent's belief in a conclusion is derived from his belief in a set of premises by means of the faculty of judgment. In light of this talk of derivation, there are a few important observations to make.

The first observation is that we seem to have a more transparent account of what it means for an agent's choice to be responsive to his best judgment. In the last chapter, I suggested that Kant takes action to be behavior that results from choice in response to best judgment, but what this responsiveness amounted to was unclear. Given everything that has been said, we can see more clearly what this responsiveness involves. Choice is responsive to best judgment in the sense that, given a best judgment, an agent responds to the best judgment by deriving a choice from this best judgment, by means of the faculty of judgment. The second observation is that Kant appears to take this relationship to be important to his analysis of rational and irrational action. Kant, after all, follows this discussion of derivation with a discussion of rational and irrational action. In light of this, there is good reason to think that our earlier account of rational and irrational action must be modified. When Kant says that rational action is action that results from one's choosing to act on reasons that are picked out by one's best judgment,

Specifically, Kant discusses the actions and angels and humans, where he discusses the role that reason has in determining the agent's choice, given his best judgment. Given the context, it seems evident that Kant thinks of this “determination” relation as being the “derivation” relation that he outlines immediately prior.
he means that rational action is action that results from one's choosing to act on reasons that are picked out by the best judgment that the choice is derived from. Likewise, when he says that irrational action is action that results from one's choosing to act on reasons that are not picked out by one's best judgment, he means that irrational action is action that results from one's choosing to act on reasons that are not picked out by the best judgment that the choice is derived from. As before, we are still interpreting Kant's statements here as giving necessary conditions for rational action and sufficient conditions for irrational action.

These observations about the relation between choice and best judgment are important, because if they are right, it appears that there is another way to draw the distinction between acting against a best judgment and merely not acting in accordance with it. After all, an agent's action could result from his choosing to act on a reason that is not picked out by some best judgment, but it might nevertheless be the case that the choice was not derived from that best judgment.51 In light of this, we might propose another account of the distinction between acting against and not acting in accordance with: One acts against a best judgment when one's action is caused by a choice to act on reasons not picked out by that best judgment and the choice in question is derived from that best judgment. One merely does not act in accordance with a best judgment when one's action is caused by a choice to act on reasons not picked out by that best judgment and the choice is not derived from that best judgment. Going back to the case of Pam, we can see how this account correctly identifies Pam as an agent who merely does not act in

51 This might initially seem like an impossible state of affairs, but attention to a variety of cases reveals that this is possible after all. Suppose, for instance, that I form a best judgment that I ought to X for reasons A, but then before acting and in light of new information, revise my best judgment to a best judgment that I ought to Y for reasons B. I then choose to Y for reasons B.
accordance with her best judgment. Remember that Pam is an agent whose action was a manifestation of a disposition – a disposition that was itself caused by a prior deliberation and choice to A for reasons \( \varphi \). In light of this, it is impossible for the choice that causes Pam's action to be derived from her current best judgment – the one that says she ought to B for reasons \( \psi \). After all, how could her past choice be derived from her future best judgment? So, according to our most recent account, Pam does not act against her best judgment, she merely does not act in accordance with it.

Keeping this second account in mind, I think that we can formulate a combined account of what it is to act against a best judgment. One acts against a best judgment when one's action results from (1) a choice to act on reasons that are not picked out by that best judgment (2) a choice to act on a reason that is contained in the domain of that best judgment, and (3) a choice that is derived from that best judgment. Now that we have this new account of what it is to act against a best judgment, we can return back to the case of Sam and see how it dissolves the first problem that I noted earlier.

**Returning Back to the case of Sam**

As noted, Sam is an agent who seemingly acts against his best judgment, and given this, the Kantian is forced to say that Sam acts irrationally. However, it now appears that the Kantian may be able to offer up a different diagnosis of Sam. According to our new account, one acts against a best judgment when one's action is caused by (1) a choice to act on a reason that is not picked out by that best judgment (2) a choice to act on a reason that is contained in the domain of that best judgment, and (3) a choice that is derived from that best judgment.
derived from that best judgment. Sam is definitely an agent whose action is caused by a choice to act on reasons not picked out by his best judgment. His action is presumably caused by a choice to study with friends for reasons $\psi$, but his best judgment says that he ought to study in isolation for reasons $\varphi$. Does Sam also satisfy the other requirements for acting against his best judgment?

It's not clear whether Sam satisfies the third requirement or not, in part because Arpaly's illustration does not provide us with enough details to determine if the choice that causes his action is derived from his current best judgment. In light of this, I suggest that we look to the second requirement for answers. Looking at this second requirement, we can see that Sam's action is caused by a choice to act on reasons that are not contained in the domain of his best judgment. This is evident because Sam's action is caused by a choice to study with friends for reasons $\psi$, and Sam does not examine reasons $\psi$ in his deliberation, so reasons $\psi$ cannot be contained in the domain of his best judgment. Since Sam's action is caused by a choice to act on reasons that are not contained in the domain of his best judgment, Sam fails this second requirement. Given this, we can conclude that Sam does not act against his best judgment after all. Moreover, since Sam does not act against his best judgment, one is not forced to claim that Sam acts irrationally.  

---

52 Of course, given that acting against one's best judgment is merely a sufficient condition for irrational action (as noted earlier), it is possible that Sam still acts irrationally. However, someone who makes this claim would have to specifically outline another condition for irrational action, and show that Sam satisfies this condition. In the absence of such an analysis, there is good reason to think that Sam doesn't act irrationally.
Some Last Thoughts

Keeping this in mind, it seems that we have a solution to the first problem offered up by Arpaly's case of Sam. The Kantian isn't forced to conclude that Sam acts irrationally, because Sam isn't an agent who acts against his best judgment in the first place. The solution was based on a closer examination of the concept of acting against a best judgment, and a closer examination of Kant's theory of practical rationality. In light of this, it seems that the project of reconstruction can achieve substantive results after all. Perhaps we should wait before we appeal to either of the simple responses given earlier.

Of course, even given this first solution, it appears that the Kantian theory of practical rationality is still in need of some further modification. The first solution assumes that Sam chooses to study with friends for reasons ψ. However, the Kantian suggests that choosing to act for a reason necessarily involves awareness of the reason that one commits oneself to act on. However, Sam does not seem to be aware of the reason that he commits himself to act on. Given this, the Kantian needs to do some work in accommodating this claim. Moreover, the Kantian still needs to deal with the second problem that I outlined earlier. Specifically, it's not clear that the Kantian theory of practical rationality can claim that Sam acts rationally. In order to claim that Sam acts rationally, the Kantian must claim that Sam acts in accordance with his best judgment. Given that Sam allegedly chooses to study with friends for reasons ψ, the Kantian must claim that Sam has a best judgment that he ought to study with friends for reasons ψ. However, Sam is not aware that he ought to study with friends for reasons ψ. Moreover, the Kantian suggests that best judgment is the product of deliberation, and Sam does not
seem to deliberate in reaching a best judgment that he ought to study with friends for reasons ψ. Given this, the Kantian theory needs to be modified in order to explain how Sam comes to this best judgment.

**Preview of the Next Chapters**

In light of the problems that remain, it is evident that the Kantian theory of practical rationality, as presented so far, is in need of some further modification. Now, it isn't immediately clear where exactly these modifications should be made, and what sorts of modifications are needed. Keeping this in mind, in the next three chapters, I would like to determine where the Kantian goes wrong, and what specific modifications are needed. In order to do this, I will examine Arpaly's own theory of practical rationality, and then examine a few theories of practical rationality that sit on the continuum between Arpaly and Kant.53

I examine Arpaly's theory of practical rationality, in part, because it accommodates the case of Sam quite well. Given this, I think it will provide us with some resources that will be beneficial to the Kantian in the end. The theories that lie on the continuum between Arpaly and Kant are theories that are much like Arpaly's own theory, but ones that incorporate different components of the Kantian theory of practical rationality. In the case that such theories accommodate the case of Sam, we will have a defense of the incorporated Kantian components. We will have shown that the Kantian theory of practical rationality does not fail to accommodate the case of Sam in virtue of

53 I should note that technically, I will be looking at Arpaly's theory of action and her theory of practical rationality. I will also be looking at theories of action and practical rationality that lie on the continuum between Arpaly and Kant. For the sake of simplicity, I will from this point refer to said theories of action and practical rationality as simply theories of practical rationality.
its incorporating these components. This will tell us that the error in the Kantian theory is to be found elsewhere. Once I find where the error is, I will propose a modified version of the Kantian theory of practical rationality. In the case that this modified theory accommodates the case of Sam, we will have shown that the Kantian can accommodate the case of Sam, given these modifications. Given that the modified theory still substantively counts as a version of the Kantian theory of practical rationality, we will have a defense of the Kantian theory of practical rationality as well. Of course, the final theory won't just provide evidence that the Kantian theory is viable. If it counts as a version of the Kantian theory of practical rationality, and moreover accommodates a substantively wider variety of phenomena in the philosophy of action, it will count as a significant improvement on the original Kantian theory presented in the first chapter.
Chapter 3: Arpaly's Theory of Practical Rationality

In this chapter, I will examine Arpaly's theory of practical rationality, as presented in her book *Merit, Meaning and Human Bondage*. The theory that Arpaly presents in this book isn't entirely complete in that there are a variety of details that need to be fleshed out. However, I think that we can reasonably reconstruct what a complete theory of practical rationality for her involves. In laying out her theory of practical rationality, I will start out by discussing two distinctions: the distinction between absolute and rationalizing reasons, and the distinction between three different kinds of cause. Given these distinctions, I will explain what Arpaly takes action to be, and moreover what she takes rational and irrational action to be. I will then show how her theory of practical rationality deals with the case of Sam, and point out why her theory succeeds.

After accomplishing this, I will discuss a Kantian criticism of Arpaly's theory of practical rationality. The criticism specifically has to do with the fact that Arpaly takes belief-desire pairs to constitute reasons for action. The Kantian will suggest that this dependence on belief-desire pairs prevents Arpaly from accounting for one way in which action involves agency. As I present it, the Kantian will suggest that Arpaly's theory cannot accommodate the intuition that action is behavior that is *guided by the agent's understanding of what he has reason to do*. As the Kantian sees it, only a theory that

---

incorporates judgments about reasons for action – that is, the first type of judgment noted in the first chapter – can account for this kind of agency. The criticism is important because it motivates our looking at a theory of practical rationality that incorporates this Kantian idea. Keeping this in mind, I will examine a new theory of practical rationality, one which is much like Arpaly's but replaces desires with what I call values. The new theory of practical rationality is one that acknowledges the aforementioned Kantian criticism, and moreover accommodates the case of Sam quite well. Given that the new theory accommodates the case of Sam, we will have a defense of a Kantian idea – that action is behavior that is guided by the agent's understanding of what he has reason to do.

After providing a defense of this Kantian concept, I will then discuss another Kantian criticism of Arpaly's theory of practical rationality. The Kantian will suggest that Arpaly's theory cannot properly account for another way in which action involves agency. Specifically, the Kantian will suggest that Arpaly's theory cannot accommodate the intuition that rational action is action that is guided by the agent's understanding of the reasons he ought to act on, or alternatively, the agent's understanding of what he has most reason to do. As the Kantian sees it, only a theory of practical rationality that employs the concept of best judgment can account for this. The criticism is important because it motivates our looking at a theory of practical rationality that incorporates the idea that action is behavior that is guided by the agent's understanding of the reasons he ought to act on. I will develop such a theory and analyze it in the next chapter.
Arpaly's Theory of practical rationality - Absolute and Rationalizing Reasons

In order to understand Arpaly's theory of practical rationality, it will help to first see how she views action. According to Arpaly, action is behavior that is *responsive to reasons*. What does she mean by “reasons” when she speaks of a behavior being responsive to reasons? Well, In *MMHB*, Arpaly introduces a distinction between “absolute” and “rationalizing” reasons that helps to clarify this. Initially, the distinction between absolute and rationalizing reasons seems to depend on whether the reason *has a bearing on the rationality of the agent in acting*. Absolute reasons appear to be reasons that *don't* have a bearing on the rationality of the agent in acting, whereas rationalizing reasons appear to be reasons that *do* have a bearing on the rationality of the agent in acting. To illustrate what this means, Arpaly presents the following pair of cases:

(3) “Before hearing the coach's view, the fact that Cox never tried long-distance swimming does not imply that anything was wrong with her as a rational or reasonable agent. She is ignorant of certain facts about long-distance swimming and the ways in which one can spot an aptitude for it, but there is nothing wrong with her reasoning. On the other hand, after she heard the coach's judgment, if she still fails to take up long-distance swimming, it is something that she has, in a sense, to account for: absent weightier reasons to the contrary, she would be irrational not to give it a try.”

(4) “If Cox took up long-distance swimming before hearing the coach's view, or without in some other way knowing of her talent, her decision would seem rash or even irrational. Why, after all, would she aim at such a crazy goal as swimming the English Channel? And at age fifteen? Absent some special story, Cox would seem like a crazy teenager who accidentally ended up doing something she had a reason to do.”

Earlier, I suggested that absolute reasons don't have a bearing on the rationality of the agent in acting, whereas rationalizing reasons do. Given the above

---

55 *MMHB*, 66.
examples we can see now what this might mean. Prior to Cox's being told of it, the fact
that Cox has an aptitude for long-distance swimming is an absolute reason for action.
The reason does not make her more *irrational* in the case that she does not proceed to
take up long-distance swimming, and moreover does not make her more *rational* in the
case that she does proceed to take up long-distance swimming. After being told of her
special aptitude, the fact that she has this aptitude becomes a rationalizing reason for
action. The reason makes her more irrational in the case that she does not proceed to take
up long-distance swimming – she would be considered irrational in not taking up
swimming, absent weightier reasons to the contrary. Moreover, the reason makes her
more rational in the case that she proceeds to take up long-distance swimming – she
would be considered rational in taking up swimming, absent weightier reasons to the
contrary.

While this way of characterizing the distinction is helpful, there are more
exact ways of marking out the distinction. On one way of looking at it, Arpaly's
distinction seems to designate a certain *class of reasons*, a class that can be divided up
into either absolute reasons or rationalizing reasons. However, given some of her later
statements, this does not seem to be an accurate way of characterizing the distinction.
From what I gather, a more appropriate way to characterize the distinction is as follows:
Absolute reasons refer to *facts* that support a certain course of action, whereas
rationalizing reasons refer to *mental states whose content* supports a certain course of
action. So, in the case of Cox, the simple fact that Cox has an aptitude for long-distance
swimming is an absolute reason for action. However, this fact does not provide a
rationalizing reason for Cox. Cox has a rationalizing reason when it's the case that she has a mental state that has the appropriate content, namely a belief that she has an aptitude for long-distance swimming. In her exposition of the distinction, it's not clear if Arpaly means to suggest that the belief gives Cox access to a rationalizing reason (access to a fact) or if the belief itself just is the rationalizing reason. However, in later passages, she treats beliefs themselves as rationalizing reasons (technically, belief-desire pairs). So from this point on, I will speak of rationalizing reasons as referring to mental states with the right sort of content – a content that supports a certain course of action.

A few more specifications will be helpful here. I just suggested that rationalizing reasons refer to mental states whose content supports a certain course of action. Keeping this in mind, it will be helpful to further specify what types of mental states she refers to when she speaks of rationalizing reasons. When she speaks of rationalizing reasons for action, she refers specifically to beliefs and desires. So, for instance, Cox has a rationalizing reason for taking up long-distance swimming when she believes that she has an aptitude for long-distance swimming, and presumably desires to take up hobbies that she has a great aptitude for. It's important to take note of this because it's easy to think that a number of other conditions are required in order for Cox to have a rationalizing reason for action. For instance, one might think that, for Arpaly, an agent has a rationalizing reason for action only when the beliefs and desires in question are conscious states of the mind. So, in the case of Cox, it might be suggested

---

56 I should point out that there is an ambiguity in Arpaly's discussion of rationalizing reasons. It's not clear if beliefs themselves are rationalizing reasons, or if belief-desire pairs are rationalizing reasons. The ambiguity isn't entirely important, but it's helpful to point out the different ways of carving up rationalizing reasons. Later on, Arpaly speaks of belief-desire pairs as being rationalizing reasons for action, so I will take this to be Arpaly's official view.
that Cox only has a rationalizing reason for action in the case that she *consciously believes* that she has an aptitude for long-distance swimming. Here, one should note that Arpaly does not take this to be a condition for having a rationalizing reason for action. For Arpaly, it is simply enough that Cox *believes* that she has an aptitude for long-distance swimming – that this belief be among the mental states that constitute her mind. This is important because *Kantians* can be construed as suggesting, as a necessary condition for having a rationalizing reason, that certain mental states be conscious states of the mind.57

It is also easy to suggest that, under Arpaly's theory, Cox has a rationalizing reason only in the case that she additionally *judges or believes* that the contents of her beliefs and desires support a certain course of action. So, in this case, Cox would have a rationalizing reason only in the case that she judged or believed that her aptitude for long-distance swimming supported her taking up long-distance swimming. It's important to note, however, that this is not Arpaly's view either. For Cox to have a rationalizing reason for action, it is enough that she simply believes that she has an aptitude for swimming and a desire to take up activities she has an aptitude for. Arpaly must commit herself to this view because Sam is an agent who simply believes that extreme isolation does not lead to success on exams, and this belief *by itself* has a bearing on Sam's rationality in acting.

57 Forgive the terminological confusion here. It can be difficult to reconcile talk of reasons between Kant and Arpaly, in part because Arpaly often refers to reasons as being *mental states* of the agent, whereas Kant tends to speak of reasons as being *considerations* that the agent takes to be normatively significant.
Three Kinds of Cause

The aforementioned distinction between rationalizing and absolute reasons is important because Arpaly takes reason-responsiveness to involve a behavior's being responsive to *rationalizing reasons*, not absolute reasons. Given this specification of what *kinds* of reasons are involved in reason-responsiveness, we can now give an initial account of what it means for a behavior to be reason-responsive. According to Arpaly, behaviors are responsive to reasons when they are *caused in the right way* by rationalizing reasons. Given moreover that rationalizing reasons consist of belief-desire pairs, we get the result that behaviors are reason-responsive when they are *caused in the right way by belief-desire pairs*. Of course, what it is for a behavior to be caused “in the right way” by belief-desire pairs is a thorny issue. In order to explain what “the right way” is, Arpaly lays out a distinction between three kinds of cause.

(1) Julius calls me an idiot, raising his voice in the process. As a result, my ears hurt.
(2) Julius calls me an idiot, raising his voice in the process. Because of a troubled upbringing and Julius's resemblance to my father, I feel diminished and inferior.
(3) Julius calls me an idiot, raising his voice in the process. As a result I conclude that Julius must be having a very hard day today.

The first kind of cause is something that we might think of as purely physical, in the sense that the *content* of Julius' utterance plays no role in the resulting event. It is evident that the content of his utterance does not play a role because Julius may well have said something else, and my ears would have hurt just the same. It is the *volume* of Julius' utterance that causes my ears to hurt, and the explanation for how high volumes

---

58 Looking at it from this perspective, Arpaly's theory of action is quite similar to Davidson's theory of action. There is a significant difference though, in that Arpaly has a more complex take on how mental states can cause behaviors in agents, as I will shortly make clear.

59 MMHB, 62.
can cause ears to hurt is a purely physical explanation, one that involves facts about the
amplitude of sound waves and the physical makeup of the pain receptors in my ear. The
second case is an example of content-efficacious cause that is not reason-responsive.
Unlike before, it seems that content-efficacy is involved because the content of Julius'
utterance and other contentful representations play a role in my reaction. If Julius had
instead told me that I was effervescent, I would not have felt diminished and inferior. If I
believed that Julius resembled a child, I might have felt annoyed instead of inferior.
However, in this second case, even though my reaction responds to the content of mental
states, Arpaly suggests that it does not involve reason-responsiveness. Why? Well, in
this second case, the contentful states that cause my reaction don't rationally support my
feeling diminished and inferior. In Arpaly's words, there is no rationalizing relationship
between the states that cause my reaction and the reaction itself.

According to Arpaly, reason-responsiveness occurs when there is content-
efficacy and the mental states that cause the behavior have a rationalizing relationship
with the behavior in question. Moreover, the mental states cause the behavior in virtue of
their rationalizing relationship with the behavior. The third case is a case that involves
both content-efficacy and true reason-responsiveness. It is content-efficacious because
the content of Julius' utterance plays a role in my reaction. If Julius had called me
effervescent, I would have likely come to a different conclusion about Julius' day.
Moreover, the mental states that cause my reaction have a rationalizing relationship with
my reaction, and they cause my reaction in virtue of their rationalizing relationship with
my reaction.
It's important to lay out all three of these kinds of cause, because Arpaly models the latter two off of the first. In the case of physical cause, Arpaly suggests that events cause one another *in virtue of their physical properties*. Here, the event in question is causally explained by mentioning a set of other events and in particular the physical properties of these events. As Arpaly sees it, the latter two types of cause can be explained in a similar manner. In the case of merely content-efficacious cause, events cause one another not in virtue of their physical properties but rather *in virtue of their content*. The explanation given for my reaction invokes the content of mental states. What about the case of genuine reason-responsive cause? Well, according to Arpaly, what marks out reason-responsive cause is that my mental states cause my reaction, *in virtue of their rationalizing relationship* with my reaction. The rationalizing relationship between my mental states and my reaction serves as part of the causal explanation of my reaction.

**What is Action?**

Earlier I noted that Arpaly takes action to involve reason-responsiveness. An action is a behavior that is *responsive to reasons*. Arpaly clarified this by asserting that reason-responsiveness involves responsiveness to *rationalizing reasons*, not absolute reasons. Rationalizing reasons specifically are constituted by *belief-desire pairs*. Given Arpaly's discussion of the three kinds of cause, we can now further clarify what it means for a behavior to be responsive to reasons. For Arpaly, a behavior is responsive to reasons when it is caused by belief-desire pairs, and these belief-desire pairs cause the
behavior in virtue of their rationalizing relationship with the behavior. Of course, since action just is behavior that is reason-responsive, we have a characterization of action as well. Action just is behavior of the agent's that is caused by belief-desire pairs, where these belief-desire pairs cause the behavior in virtue of their rationalizing relationship with the behavior. Actually, Arpaly is a bit more specific: an action is a behavior that is immediately caused by belief-desire pairs, where the belief-desire pairs cause the behavior in virtue of their rationalizing relationship with the behavior. Mere behavior occurs when the behavior is immediately caused by a non-rationalizing cause.

**What is Rational Action?**

As noted, action just is behavior that is immediately caused by belief-desire pairs, where the belief-desire pairs cause the action in virtue of their rationalizing relationship with the action. Now that we have a sense of what action is, we can turn to an account of what rational and irrational action are. What is rational action? Well, in order to explain what rational action is, we first have to make sense of an analysis of rational action that Arpaly commits herself to. According to Arpaly, rational action is action that is done *for the best reasons* and moreover done for these reasons *because they are the best reasons*. Keeping this in mind, in order to present her account of rational and irrational action, we'll have to explain how her theory accommodates this analysis.

What are the agent's “best reasons”? Unfortunately, Arpaly doesn't give an

---

60 I should note that Arpaly *never* outright provides this analysis. She discusses agents who act rationally as being agents who act for *good reasons* – where she means to say that the agent acts on their particular reasons *because* they are good reasons. So, the above analysis is a reasonable interpretation of what she says in her written work. Also, I have confirmed in personal communication that Arpaly commits herself to the above analysis that I provide on her behalf.
account of what the agent's best reasons are, but we can construct an account that is consistent with her statements, given a few assumptions. As noted earlier, belief-desire pairs have rationalizing relationships with actions, which is just to say that they rationally support various courses of action. Keeping this in mind, we might think that at any given time, an agent's body of belief-desire pairs rationalize (or rationally support) a variety of different possible actions. Given this set of possible actions, we might also think that some of these actions are *better rationalized* by the agent's belief-desire pairs than others, and that some actions are *most rationalized* by the agent's belief-desire pairs as well. In a given situation, the agent's body of belief-desire pairs makes it the case that it is *more rational* to X than to Y, and that it is *most rational* to Z. Arpaly isn't clear on exactly what makes it the case that an action is most rationalized by the agent's belief-desire pairs, but a number of different stories are compatible with her account. For the time being, let's call the action that is most rationalized by the agent's belief-desire pairs the “best action.”

Keeping this in mind, we might then have the following account of the agent's “best reasons.” The agent's “best reasons” are those reasons (I.e., belief-desire pairs) that support the agent's “best action.” This characterization of “best reasons” does give us the first part of Arpaly's analysis of rational action. Rational action is, in part, action that is done *for the best reasons*. Given Arpaly's account of how belief-desire pairs cause action, we can see that for Arpaly, an action is done *for the best reasons* when the action is immediately caused by the belief-desire pairs that represent the agent's best reasons, where the belief-desire pairs cause the action in virtue of their rationalizing
relationship with the action.

Now that we have this account, we can turn to the second part of Arpaly's analysis. What is it for the agent to act for the best reasons *because they are the best reasons*? This is a bit more tricky to figure out, but I think if we take on a few basic assumptions, we can see how she explains this. Let's assume, for instance, that a belief-desire pair exerts a *rationalizing causal force* that corresponds directly to the strength of the belief-desire pair as a reason for action. Let us also assume that for a *set* of belief-desire pairs that support a particular action, the rationalizing causal force exerted by that set of belief-desire pairs corresponds directly to the *cumulative* strength of the rationalizing causal forces exerted by the belief-desire pairs in that set. In this case, we might think that the belief-desire pairs that constitute the agent's best reasons *naturally* exert the greatest rationalizing causal force on the agent's behavior in comparison with other belief-desire pairs. Finally, let us assume that the set of belief-desire pairs that exerts the greatest rationalizing causal force ends up being the set of belief-desire pairs that immediately causes the agent to act, in virtue of their rationalizing relationship with the action.

Keeping these assumptions in mind, we might think that in the normal course of events, an agent will act for his best reasons because his best reasons naturally exert the greatest rationalizing causal force on his behavior. Moreover, these best reasons will exert the greatest rationalizing causal force on his behavior *because they are the best reason*. Given this, we can literally say that the agent acts for the best reasons *because they are the best reasons*. So, we now have an analysis of what it means to act for the
best reasons because they are the best reasons, and given this we also have our analysis of rational action as well. As Arpaly suggests, rational action is action that is done for the best reasons, and moreover done for these reasons because they are the best reasons. This happens when the agent's behavior is immediately caused by belief-desire pairs that represent the agent's best reasons, where these belief-desire pairs cause the action in virtue of their rationalizing relationship with the action. Moreover, the belief-desire pairs that represent the agent's best reasons cause the agent's action, because the best reasons naturally exert the greatest rationalizing causal force, in virtue of the fact that they are the best reasons.

**What is Irrational Action?**

Now that we have an account of rational action, we can turn to an account of irrational action. As I interpret it, Arpaly sees irrational action as simply being action that is done *for reasons other than the best reasons*. Given our previous discussion, we can cash this out as follows: Irrational action is action that is immediately caused by belief-desire pairs that do not represent the agent's best reasons, where these belief-desire pairs cause the action in virtue of their rationalizing relationship with the action.

This account is pretty simple, but one might wonder exactly *how* it is that an agent acts for reasons that are other than his best reasons in the first place. Earlier, I noted that belief-desire pairs exert a rationalizing causal force that corresponds to their strength as reasons, and as a result, the belief-desire pairs that represent the agent's best reasons exert the greatest rationalizing causal force. In light of this, the agent naturally
acts on his best reasons. Keeping this in mind, it's not clear how an agent could act on reasons that are other than his best reasons. Regarding how irrational action might happen, Arpaly makes the following observation: “On my view, human beings being what they are, the best overall explanation of a piece of human behavior or belief will often be an explanation that presents it as a mixture of the reason-responsive, the merely-content-responsive, and the robot-like – and woven fine, too.”\textsuperscript{61} As we can see here, Arpaly suggests that an action may have a number of different causes, and giving a complete causal explanation of an action may involve pointing to a number of different causal factors. Given this, it is oftentimes the case that causal factors other than the agent's rationalizing reasons affect the behavior of the agent.

How do they affect the behavior of the agent? Arpaly suggests that non-rationalizing causes play a role in the agent's action as secondary causes, inhibiting or enhancing the rationalizing causal force of the agent's belief-desire pairs, thus making it the case that the agent acts on reasons that are other than the best reasons. These non-rationalizing causes can, for instance, enhance the rationalizing causal force of reasons that are not the best reasons, or they can inhibit the rationalizing causal force of the agent's best reasons. Of course, sometimes these non-rationalizing causes can help the agent to act on his best reasons as well. Arpaly notes that “the atavistic factor only makes one irrational if it prevents one from acting for some good rationalizing reasons. Some atavistic factors do not do that: in fact, they fight those factors that would prevent one from acting for good reasons, and in this way enhance the efficacy of one's rationalizing

\textsuperscript{61} MMHB, 72.
So, non-rationalizing causes can influence whether the agent acts on his best reasons or other reasons. That said, although such non-rationalizing causes can affect the causal efficacy of the agent's rationalizing reasons, they do nothing to change the bearing that such rationalizing reasons have on the agent's rationality in acting. They may cause the agent to act on reasons that are other than the best reasons, but they do not make these other reasons the best reasons for action.

**Returning Back to the Case of Sam**

Now that we finally have a more or less complete version of Arpaly's theory of practical rationality, we can discuss how Arpaly's theory of practical rationality accommodates the case of Sam. As it turns out, Arpaly's theory deals with the case of Sam quite well. Let me briefly restate the case of Sam again: Sam is a college student who has previously tried to pass his exams by isolating himself from friends and studying feverishly. Unfortunately, this course of action has never led Sam to be particularly productive. Instead of being productive, he gets lonely and plays video games. The result is that Sam never does very well on his exams when he isolates himself from the outside world. Sam comes to believe that his isolation never gets him the results that he wants, but he never reflects on this fact in any normatively significant way – he never notices how this fact might present him with a reason against extreme isolation.

Before the end of the next semester, Sam notices his exams are coming up sooner than expected and this makes him anxious. He is aware of a variety of reasons for action, but he fails to remember that extreme isolation causes him to do badly on his exams.  

62 MMHB, 76.
exams. As a result, when he deliberates about what he ought to do, he comes to the conclusion that he ought to study in isolation again. However, Sam does not in fact proceed to isolate himself. Instead, Sam finds himself studying with friends – a tactic that does not make him so isolated, but nevertheless makes him significantly more productive. Moreover, Arpaly states that unbeknownst to him, Sam acts for the best reasons, among them the reason that he failed to attend to in his deliberation. In fact, according to Arpaly, Sam acts on such reasons because they are the best reasons that he has for action. In other words, Sam acts rationally.

So how does Arpaly account for this case? Well, rational action occurs when an agent acts for the best reasons, and moreover acts on these reasons because they are the best reasons. So, in order to accommodate the case of Sam, Arpaly just needs to explain how Sam does this. Arpaly's analysis is simple. According to Arpaly, Sam's entire belief-desire set is such that it rationally supports his studying with friends. This tactic is most rationalized given the belief-desire set that he has. In light of this, the best reasons that Sam has for action are reasons that support his studying with friends (these are reasons $\psi$). Given that Sam does in fact act for these reasons, Sam acts for the best reasons. More specifically, the belief-desire pairs that constitute these reasons immediately cause him to act, in virtue of their rationalizing relationship with his action. Sam, of course, is not aware of these reasons in acting, but this is compatible with Arpaly's theory. Belief-desire pairs exert a rationalizing causal force on action purely in virtue of the fact that a rationalizing relationship exists between the belief-desire pairs and the action. In light of this, Arpaly is free to say that such belief-desire pairs cause
him to act, even though he isn't aware of them. Moreover, Arpaly can explain how Sam acts on these reasons \textit{because they are the best reasons}. Since his best reasons exert the greatest rationalizing causal force, they naturally end up being the reasons that cause him to act as well. Given that they exert the greatest rationalizing causal force in virtue of the fact that they are the best reasons, we get the result that Sam acts on his best reasons \textit{because they are the best reasons}. Since Arpaly has a story explaining how Sam acts for the best reasons, and acts for these reasons because they are the best reasons, Arpaly can suggest that Sam acts rationally. So, Arpaly's theory of practical rationality successfully accommodates the case of Sam.

At this point, it's instructive to see how Arpaly's theory of practical rationality holds up against the Kantian theory, and see why it is able to deal with the case of Sam so well. The most obvious benefit that we can see comes from the fact that Arpaly \textit{does not employ the concept of best judgment} in her theory of practical rationality. In light of this, she can easily secure the conclusion that Sam acts rationally even though he seemingly acts against his best judgment. It is simply enough that he acts for the best reasons and moreover acts on these reasons because they are the best reasons. Since Arpaly's theory of practical rationality can accommodate this analysis of rational action, she is able to explain how Sam acts rationally. Since her analysis of rational action does not employ the concept of best judgment, Sam can literally have \textit{any} best judgment and it still could be the case that Sam acts rationally. Of course, the Kantian does have a partial response. As I suggested in the previous chapter, the Kantian can bypass the conclusion that Sam acts against his best judgment, and as a result acts irrationally. However, it seems that
the Kantian cannot explain how Sam acts rationally. The fact that Arpaly can account for how Sam acts rationally suggests that Arpaly is on a better footing.

Consider that in order for the Kantian to say that Sam acts rationally, he has to suggest that Sam has a best judgment telling him he ought to study with friends for reasons $\psi$. Moreover, in order to say that Sam has this best judgment, the Kantian must say that Sam \textit{deliberated} in coming to this best judgment. However, Arpaly makes it evident that Sam has not deliberated \textit{at any time} in coming to this best judgment.

Consider also that Arpaly's theory of practical rationality does \textit{not} employ the concept of deliberation. In light of this, her theory doesn't run into the pesky awareness requirements associated with the Kantian notion of deliberation. All of the work in Arpaly's theory is done by belief-desire pairs and the rationalizing relationships that exist between actions and belief-desire pairs. Awareness is not required at all for her theory, and so her theory does not suffer from the problems associated with the Kantian theory.

\textbf{Developing a New Theory of Action}

In light of what has been said, it appears that the Kantian theory of practical rationality is in need of some work. The Kantian theory of practical rationality cannot accommodate the case of Sam without some changes. Arpaly's theory of practical rationality, on the other hand, seems to deal with the case of Sam quite well. Keeping this in mind, I will now turn to the project that I outlined at the end of the last chapter and the beginning of this chapter. I will present a Kantian criticism of Arpaly's theory of practical rationality, and show how it motivates the addition of a component from the
Kantian theory of practical rationality. I will present a new theory of practical rationality, one which incorporates the Kantian component, and then show how this theory of practical rationality accommodates the case of Sam. In the case that the new theory accommodates the case of Sam, we can conclude that the Kantian theory of practical rationality does not fail in virtue of the incorporated component. This process will show where the Kantian theory of practical rationality doesn't need to be modified, and will eventually show where a modification is needed.

The First Kantian Criticism

Let me begin with a standard criticism that Kantians have presented for theories like the one that Arpaly champions. The problem specifically has to do with her theory's dependence on beliefs and desires. According to this standard criticism, beliefs and desires essentially do all of the work in her theory, since they constitute reasons for action and they cause actions, in virtue of their rationalizing relationships with actions. Keeping this in mind, Kantians will object to Arpaly's theory because it leaves the agent completely out of the theory. According to the Kantian, beliefs and desires are entities with respect to which agents are passive, and if beliefs and desires do all of the work in her theory, then her theory is one in which agents are absent.63 How do Kantians incorporate the agent into their theory of practical rationality? Well one way they do this is by having the agent play a role at the level of reasons. Consider that for the Kantian, an agent doesn't have a reason for action (i.e., a “rationalizing reason”, in Arpaly's terms)

63 For an example of this type of criticism, see Barbara Herman in The Practice of Moral Judgment, Korsgaard in The Sources of Normativity and Wallace in “Three Conceptions of Rational Agency”.

84
until he judges a maxim to pass an imperative test – he judges that a certain consideration is a reason for action. For the Kantian, it is these judgments that provide an agent with reasons for action. Moreover, according to the Kantian, judgments about reasons for action are entities with which agents are active, and if these judgments provide us with reasons for action, then agents play a role in action by playing a role at the level of reasons.

Now, while this standard Kantian criticism is initially appealing, it isn't entirely unproblematic. In light of this, let me spend some time exploring the criticism further. For starters, we can note that Arpaly has her own explanation for how her theory accounts for the agency in action. According to Arpaly, there is an uncontroversial sense in which beliefs and desires belong to the agent - they are the agent's mental states. In light of this, there is a sense in which Arpaly's theory of practical rationality incorporates the agent after all. Her theory incorporates the agent by having mental states of the agent (i.e., the agent's beliefs and desires) play a causal role in bringing about action. Of course, a Kantian might look at Arpaly's story here and make a criticism that is similar to the earlier one: Although Arpaly's theory has it that the mental states of the agent play a causal role in bringing about action, not all mental states of the agent can cause action, as opposed to mere behavior. After all, there is a sense in which a gut reflex is a mental state of the agent, and yet the behavior that the gut reflex causes should not be called action. The problem is to find out which mental states that belong to the agent are representative of the agent in action, and which mental states are not. Here, the Kantian can provide an appealing distinction that helps to determine what counts as being
representative of the agent in action – the active/passive distinction. Only mental states with respect to which the agent is active can truly represent the agent in action. Given that beliefs and desires are not states with which an agent is active, such mental states cannot represent the agent in action, and hence Arpaly still cannot account for the agency in action.

At this point though, Arpaly can offer up an obvious response: It's not entirely clear that agents are actually active with regards to their judgments about reasons for action, or, for that matter, passive with regards to their beliefs and desires. Aside from a mere stipulation on the part of the Kantian (and oftentimes it looks like just that), it's hard to provide some independent standard that determines whether an agent is active with regards to one and not the other. Moreover, given that the Kantian cannot adequately distinguish between states based on activity and passivity, this puts into question the legitimacy of the distinction in the first place. After all, what good are distinctions if one cannot make distinctions with them? This, I think, is a good response to make on behalf of Arpaly, and I think it reveals a problem with employing the active/passive distinction in arguing for the Kantian theory of practical rationality. In light of this, I think we can make a bit more headway if we scrap the active/passive distinction and look back at what it is that makes judgments about reasons for action so special. As I see it, these judgments are special in two respects. First, they have as their content the idea that a certain set of considerations are reasons for action. Second, they have a necessary connection to the awareness of the agent – such judgments are essentially a special kind of organized representation that is necessarily accompanied by
self-consciousness. Keeping this in mind, why is it that these judgments about reasons for action are necessary in accounting for the agency in action? As I see it, these judgments secure a sense in which an agent's understanding of what he has reason to do genuinely guides his action. Given their content – that a certain set of considerations are reasons for action – they count as something that represents an understanding of what one has reason to do. Given their necessary connection to the agent's awareness, they count as something that is truly representative of the agent. Keeping this in mind, we have a refined criticism to present to Arpaly's theory of practical rationality. Given that Arpaly's theory does not make use of anything like a judgment about reasons for action, Arpaly's theory cannot accommodate this type of understanding and so cannot account for one way in which action involves agency.

Why is it that one would take it to be the case that action is behavior that is genuinely guided by the agent's understanding of what he has reason to do, in the first place? I think that this Kantian claim is motivated by two sub-claims. The first sub-claim is that a look at conscious deliberate action shows that agents who perform such actions take their actions to be genuinely guided by this understanding. The idea is that when an agent performs a conscious deliberate action, there is a significant sense in which he takes his action to be partially guided by his understanding that certain considerations are reasons for action. After all, it is the agent's understanding that certain considerations are reasons for action that allows him to then determine what reasons he

---

64 Kant has an admittedly controversial account of what it is for a representation to be a conscious state of an agent. For him, a representation is a conscious state of the agent when the representation is accompanied by the “I think” of self-consciousness. It should be noted, however, that nothing in the dissertation hinges on this specific claim made by Kant. All that we need to note is that Kant seems to take it as a requirement that the representation be a conscious state of the agent.
ought to act on, and hence what reasons he commits himself to act on. Similarly, there is also a significant sense in which such an agent cannot deliberately act unless he first comes to an understanding of what he has reason to do in the first place. The second sub-claim is that conscious deliberate action is to be thought of as the paradigm case of action. The idea here is that if we are to build a theory of action, conscious deliberate action ought to be the type of action that serves as the model for action in general. While I think that the first of these two sub-claims is relatively uncontroversial (in fact, I think many Humeans may readily grant it), the second sub-claim has been the subject of some speculation – in particular, Arpaly has herself attacked this second sub-claim on a number of occasions. Regarding this speculation, let me point out a significant reason why I think Kantians take conscious deliberate action to be the paradigm case of action. As Kantians see it, conscious deliberate actions are actions that we take to be paradigm cases of agency in action – agents are most clearly involved in action, and are most clearly held responsible for their actions when they act in a consciously deliberate manner. Given that agency is taken to be a definitive characteristic of action, it seems that there is good reason then to think that conscious deliberate action is the paradigm case of action.

Getting back to my earlier criticism of Arpaly's theory, I should note that the criticism I am suggesting is not to be taken as employing the active/passive distinction. I am not, for instance, suggesting that the agent's activity in action is manifested in the role that the agent's understanding of what he has reason to do plays in action. Rather, I

65 We can see a claim similar to this one being made by Korsgaard in the Sources of Normativity. Korsgaard, C. The Sources of Normativity. Cambridge University Press, Cambridge, 1996. Henceforth “SN”.

66 Perhaps there is a way of construing this criticism as employing a sophisticated interpretation of the active/passive distinction, but for the purposes of this dissertation, I do not mean to be construing the criticism in this way.
am suggesting that Kantians reject the active/passive distinction as a useful distinction in arguing against Arpaly. Instead, I think that one should account for the agency in action by showing that the agent's understanding of what he has reason to do plays a role in his action – one shows that this understanding plays a role in his action by showing that the agent's judgments about reasons for action play a role in his action. Keeping this criticism in mind, Arpaly could suggest that her own theory can in fact account for the agent's understanding of what he has reason to do. I will not claim that Arpaly cannot do this. However, let me point out two difficulties that make accomplishing this task difficult: (1) The natural way to account for an agent's understanding of what he has reason to do is to say that the agent understands that X, and the natural way to account for this is to point to the agent's possession of a mental state that has X as its content. Given that Arpaly's theory does not employ a mental state that has as its content the idea that a consideration is a reason for action, Arpaly would have to take on some controversial assumptions regarding what understanding amounts to. (2) Supposing that Arpaly could come up with a plausible story under which her theory accounts for the agent's understanding of what he has reason to do, given that under her theory, beliefs and desires by themselves already entirely determine action, it's not clear that she can also claim that the agent's understanding of what he has reason to do genuinely guides his action. This is to say that any attempt at accounting for this type of understanding would leave such an understanding as an epiphenomenal byproduct – an understanding that exists but is genuinely causally unefficacious in action. If this is right, then it's not clear that Arpaly can really account for the idea that such an understanding
Returning back to the original criticism, I simply mean to point out that there are significant considerations that motivate the incorporation of judgments about reasons for action into a theory of practical rationality. If one thinks that it is important, in accounting for the agency in action, to accommodate the idea that an agent's understanding of what he has reason to do genuinely guides his action, then there is good reason to think that judgments about reasons for action ought to be included in a theory of practical rationality. Keeping this in mind, in the next section, I will examine a theory of practical rationality which is much like Arpaly's theory, except that it accommodates this Kantian idea.

The New Theory of Action

Given the clarification of the Kantian criticism, we can understand the Kantian motivation for incorporating judgments about reasons for action into a theory of practical rationality. In light of this, we can now return to our earlier project and see if we can develop a theory of practical rationality that incorporates this idea, while still accommodating the case of Sam. If this is possible, then we will have a kind of defense of a specifically Kantian idea – that action proper requires that an agent's understanding of what he has reason to do genuinely guides his action. Essentially, if a theory that incorporates this idea can accommodate the case of Sam, then we will have shown that if there is a problem with the Kantian theory of practical rationality (as outlined in the first

---

67 Perhaps Arpaly will simply reject that action is genuinely guided by an agent's understanding of what he has reason to do because it over-intellectualizes action? Since a similar charge by Arpaly can be brought up throughout the dissertation, I will address this charge at the end of the fifth chapter.
chapter), then it is not in virtue of its incorporation of this idea.

What would such a theory of practical rationality look like? Well, we do know that Arpaly's own theory of practical rationality accommodates the case of Sam, so perhaps we should start with Arpaly's theory and supplement it with the Kantian notion of judgment. As noted, Arpaly's theory begins with her contention that belief-desire pairs constitute reasons for action. Perhaps we can modify Arpaly's theory of practical rationality and create a new theory by replacing desires with judgments. Specifically, we can suggest that belief-judgment pairs constitute reasons for action. What sorts of judgments do I have in mind? Here, I refer to the judgments that result from the agent's employing the faculty of judgment and imperatives in the evaluation of individual maxims (i.e., the first type of judgment). If Herman is correct, these judgments are retained by the agent in his conception of the good after he forms them. In light of this, we can expect that an agent will have a conception of the good that is populated by a large number of these judgments. Moreover, these judgments essentially make claims about whether one ought to act on a particular maxim – they say roughly that a certain type of consideration constitutes a reason for a certain type of action. In light of this, we can see how these judgments can provide specific reasons for action, given that one has beliefs of the right sort. My judgment that hunger constitutes a reason for eating, and my belief that I am hungry plausibly constitute a reason for eating.68

68 In case one thinks this is an odd suggestion, one should note that Herman contends that such judgments combine with the agent's perception of his situation and principles of salience to make the agent aware of a reason to act. This suggestion is not all that different. Regarding the content of judgments, one should be careful to distinguish the judgment that hunger constitutes a reason for eating, from the judgment that eating will drive away hunger. The first is a judgment with normative content which identifies a certain set of considerations (hunger) as a reason for an action. The second is merely a theoretical judgment which identifies the action as being a causally sufficient means to achieving an end. I am speaking of judgments in the first sense – a judgment which specifies a certain type of
In order to avoid any confusion with later claims, let me refer to these aforementioned judgments – that is the judgments about reasons for action that are retained by the agent in his conception of the good – as values. Keeping this in mind, the new theory of practical rationality is one that claims that belief-value pairs constitute reasons for action. Since values have as their content the idea that certain considerations are reasons for action, they secure the sense in which an understanding of what one has reason to do guides the agent's action. Moreover, since values have a connection to the agent's awareness (that is, the agent is necessarily aware in the formation of values), they count as being representative of the agent in action. Let me fill in some details. What is action, according to this new theory? Following Arpaly's lead, we can suggest that action occurs when an agent's behavior is immediately caused by belief-value pairs, where the belief-value pairs cause the agent's action in virtue of their rationalizing relationship with the action.69

What are rational and irrational action, according to this new theory? Well, we might follow Arpaly's lead again. As suggested earlier, rational action is action that is done for the best reasons and moreover done for these reasons because they are the best reasons. If “best reasons” are taken to be reasons that support the “best action,” then we already know what it is for an agent to act for the best reasons. An agent acts for the best reasons when the belief-value pairs that represent the agent's best reasons immediately cause the agent's action, where these belief-value pairs cause the action in virtue of their consideration as being a reason for a certain type of action. My suggestion above is essentially that this judgment, in conjunction with the belief that the considerations in question obtain, provides the agent with a reason for action.

69 One might wonder exactly how this view differs from Arpaly's – we should note here that the difference mainly lies in the difference between desires and values. A value (and not a desire) incorporates the agent's own evaluation of the reason-giving status of his beliefs and desires, whereas a desire does not.
rationalizing relationship with the action. Moreover, if we assume that belief-value pairs exert a rationalizing causal force that corresponds to their strength as reasons, then it will naturally be the case that an agent's best reasons immediately cause the agent's action, since they (as a whole) exert the greatest rationalizing causal force on the agent's action. An agent acts for the best reasons because they are the best reasons in the sense that the agent's best reasons cause him to act, due to their having the greatest rationalizing causal force, a rationalizing causal force that they have in virtue of their being the best reasons.

We also have an account of irrational action. Irrational action just is action that is done for reasons that are not the best reasons – action that is immediately caused by belief-value pairs that do not represent the agent's best reasons, where these belief-value pairs cause the agent's action in virtue of their rationalizing relationship with the action. How is it that an agent's action can be caused by such belief-value pairs in the first place? Well, the story would again be similar to Arpaly's. Even though the agent's best reasons naturally have the greatest rationalizing causal force, other factors can play a role as secondary causes and inhibit or enhance the causal efficacy of the agent's belief-value pairs. If such factors sufficiently inhibit the causal efficacy of the agent's best reasons or sufficiently enhance the causal efficacy of reasons that are not the best reasons, then reasons that are not the best reasons will play the role of immediate cause. In such a case, the belief-value pairs that have the greatest rationalizing strength will not have the greatest causal efficacy, and irrational action will result.
How This Account Deals with Sam

Since this theory of rationality is similar to Arpaly's, we can expect it to explain the case of Sam in a similar manner. According to this theory, Sam's entire belief-value set is such that it most rationalizes his studying with friends. Given this, the best reasons that Sam has for action are reasons that support his studying with his friends (i.e., reasons \( \psi \)). Since Sam does in fact act for these reasons, it is evident that Sam acts for the best reasons. Specifically, the belief-value pairs that represent his best reasons immediately cause his action, where these belief-value pairs cause his action in virtue of their rationalizing relationship with his action. The theory can also explain how Sam acts on these reasons because they are the best reasons. Since these reasons are the best reasons, they naturally exert the greatest rationalizing causal force on Sam's behavior, and thus become the reasons that immediately cause him to act. His best reasons cause him to act because they exert the greatest rationalizing causal force, a rationalizing causal force that they have because they are the best reasons. In light of this, it literally is the case that Sam acts for the best reasons because they are the best reasons.

At this point, we should observe that this theory of practical rationality succeeds in accommodating the case of Sam for the very same reasons that Arpaly's theory succeeds. Since the theory does not employ the concept of best judgment, it can suggest that Sam acts rationally even though he acts against his best judgment. All that is required of Sam is that he act for the best reasons and moreover act on these reasons because they are the best reasons. In addition to this, since the theory does not employ the concept of best judgment, it also does not need to employ the concept of deliberation.
either. Given this, the theory does not need to suggest that Sam deliberates in any way prior to his action. It thus avoids the complications that the Kantian theory suffered as a result of its deliberation requirement. Finally, since Sam's belief-value pairs cause his action in virtue of a rationalizing relationship that exists between them, Sam does not need to be aware of anything in order for his belief-value pairs to cause his action. In light of this, the theory avoids all of the pesky awareness requirements associated with the original Kantian theory of practical rationality.

**A Potential Problem**

There is one potential problem with this new theory, and this has to do with the fact that awareness is involved in the formation of values. As I noted earlier, values are essentially judgments that result from the agent's employing the faculty of judgment and imperatives in the evaluation of individual maxims. In light of this, it is evident that awareness is required for the formation of values. Specifically, the formation of a value requires that the agent be aware of a consideration, and through judgment and the imperatives, come to be aware of a reason for action. Of course, once the agent forms the value, he retains it in his conception of the good, whether he is aware of it or not. Moreover, once the value is retained in the agent's conception of the good, the agent does not need to be aware of the value in order for it to play a causal role in his action. Belief-value pairs cause action in virtue of the rationalizing relationship that exists between them, whether the agent is aware of the value or not. However, the formation of a value does involve the awareness of the agent. The question is whether or not this creates a
problem in dealing with the case of Sam.

I think it does not. In order to see why, it will help to take a look at the exact values that constitute Sam's reasons for studying with his friends. Arpaly's story does not make it immediately clear what sorts of values would be relevant, but we can reasonably reconstruct what they might be. Sam, for instance, judges that success on exams is a consideration that constitutes a reason for studying with friends, and he judges that the development of friendships is a consideration that constitutes a reason for studying with friends. In light of this, we now have to ask ourselves: Is it psychologically plausible to suppose that Sam has formed these values in the past? I think it would be implausible to say otherwise. We would have to imagine that for over a decade, Sam never consciously reflected on the relationship between academic success and studying with friends, or never reflected on the relationship between developing friendships and studying with friends.

As unreflective as we would like to think today's college students are, it would be odd to think that Sam never makes judgments like these. All he needs is a call from a friend, asking “Hey, do you want to study for the exam with me today?” and this would get Sam to form the appropriate judgments. A teacher's assigning a group project would get Sam to form the appropriate judgments. The simple observation that his friend, Julie, got a better grade than he did because she studied with “the smart kid” would get Sam to form the appropriate judgments. The point is that situations that elicit judgments like these happen all the time. Of course, perhaps Sam was home-schooled and grew up in a Skinner Box, but this possibility aside, it's hard to see how he would not
form these judgments in the past. After all, Sam is a normal everyday college student. If this is right, then the theory of rationality I have offered up avoids the aforementioned problem.70

The Second Kantian Criticism

So far, I have discussed Arpaly's theory of practical rationality, and laid out a Kantian inspired version of Arpaly's theory that replaces the desires in Arpaly's theory with values. Given that this new theory of practical rationality successfully accommodates the case of Sam, we have a defense of the idea that action proper requires that an agent's understanding of what he has reason to do genuinely guides his action. At this point, we can now turn to another Kantian criticism of Arpaly's theory of practical rationality, a criticism that will motivate the addition of another Kantian concept.

To get at the problem, we can start out by observing that the Kantian theory of practical rationality is also able to accommodate the analysis of rational action presented by Arpaly – that rational action is action that is done for the best reasons and moreover done for these reasons because they are the best reasons. How does the Kantian do this? Well, the Kantian theory interprets both parts of the analysis using the concept of best judgment. According to the Kantian, an agent acts rationally when his action results from his choosing to act on the reasons picked out by his best judgment (he acts in accordance with his best judgment). The Kantian accommodates the first half of

---

70 We have to also remember that judgment is not an overwhelmingly taxing affair. While critics of Kant oftentimes describe the process of judgment as being a deliberate step-by-step process that involves consciously reflecting on imperatives, Kant himself does not see it in this way. As noted in the first chapter, judgment can simply involve being aware of considerations, and coming to be aware that the considerations constitute a reason for performing the action given in the maxim.
the analysis by taking “the best reasons” to refer to the reasons picked out by the agent's best judgment. Keeping this in mind, an agent whose action results from his choosing to act on the reasons picked out by his best judgment does in fact act for the best reasons. The Kantian accommodates the second half of this analysis by pointing to the role that best judgment plays in getting the agent to act on his best reasons. According to the Kantian, an agent who acts rationally chooses to act on the reasons that he does because his judgment tells him they are the best reasons for action – that they are the reasons he ought to act on. Remember that in the second chapter, it was revealed that Kant takes the agent's choice to be derived from his best judgment in a significant way. This is just to say that for the Kantian, an agent who acts rationally really does take his best judgment to play a significant role in determining what reasons he chooses to act on. Keeping this in mind, there is an obvious sense in which the Kantian accommodates the idea that an agent who acts rationally acts on his best reasons because they are the best reasons.

Now that we see how the Kantian accommodates the earlier analysis of rational action, we can turn to the criticism outright. The criticism is similar to one that I presented earlier on in this chapter. Arpaly's theory fails to capture another important sense in which agents must be involved in their actions. Specifically, the Kantian will suggest that an adequate theory of practical rationality should be able to capture the sense in which rational action is guided by the agent's understanding of the reasons he ought to act on – or the agent's understanding of what he has most reason to do.\textsuperscript{71} (Here, we should be careful to distinguish the agent's understanding of what he has reason to do

from the agent's understanding of what he has *most reason* to do). Arpaly fails to capture this sense of agency because of the way in which she accommodates the analysis of rational action given earlier. Arpaly defines “best reasons” in a way that is *independent* of the agent's understanding of the reasons he ought to act on. Moreover, she accounts for how rational agents act for the best reasons *because they are the best reasons* by pointing to the *rationalizing causal force* that such reasons exert on the agent's behavior. The agent naturally acts on the best reasons because the best reasons exert the greatest rationalizing causal force, a rationalizing causal force that they have *because* they are the best reasons. Keeping this in mind, it is evident that the agent's understanding of the reasons he ought to act on is irrelevant to Arpaly's theory. If this is right, then the agent's understanding of the reasons he ought to act on does not guide his action, and so Arpaly fails to capture another sense in which action involves agency.

Why is it that one would take it to be the case that rational action is behavior that is genuinely guided by the agent's understanding of the reasons he ought to act on? As before, I think that this Kantian claim is motivated by two other claims: that agents engaged in conscious deliberate action take themselves to be guided by their understanding of what reasons they ought to act on, and that conscious deliberate action is the paradigm case of action. Regarding the first claim, the idea is that when we look at examples of conscious deliberate action, agents who perform such actions take their actions to be guided by their understanding of what they have most reason to do. Regarding the second claim, the idea is that conscious deliberate action ought to be considered the paradigm case of action because *agency* is taken to be a definitive
characteristic of action, and actions most clearly express agency when it is the case that they are consciously deliberate actions.

In response to my criticism, Arpaly could suggest that her view can account for the agent's understanding of the reasons he ought to act on. Again, I will not argue that Arpaly cannot do this, but let me point out some difficulties for Arpaly: (1) The natural way to account for an agent's understanding of the reasons he ought to act on is to say that the agent understands that X, and the natural way to account for this is to point to the agent's possession of a mental state that has X as its content. Given that Arpaly's theory does not employ a mental state that has as its content the idea that certain reasons are the reasons one ought to act on, Arpaly would have to take on some controversial assumptions regarding what this understanding involves. (2) Given that for Arpaly, an agent's acting for the best reasons “because they are the best reasons” is entirely determined by the agent's beliefs and desires, it's not clear that she can also claim that the agent's action is genuinely guided by the agent's understanding of the reasons he ought to act on. The agent's understanding of the reasons he ought to act on can, at best, be an epiphenomenal byproduct – an understanding that exists but is genuinely causally unefficacious in action. If this is right, then it's not clear that Arpaly can really account for the idea that such an understanding genuinely guides the agent in action.2

The Kantian theory of practical rationality, on the other hand, does make sense of the idea that rational action is genuinely guided by the agent's understanding of the reasons he ought to act on. The Kantian defines “best reasons” as being the reasons picked out by

---

2 Perhaps Arpaly will suggest that the latest Kantian criticism over-intellectualizes action in a way that is objectionable. Again, as this charge comes up several times in the dissertation, I will deal with it at the end of the fifth chapter.
the agent's best judgment. Moreover, the Kantian takes rational action to be action that is 
guided by the agent's best judgment. Since the agent's best judgment represents the 
agent's understanding of the reasons he ought to act on, the agent's understanding of the 
reasons he ought to act on genuinely guides rational action, according to to the Kantian 
theory. Keeping this in mind, we can see that only a theory that incorporates best 
judgment can bypass the criticism that the Kantian makes against Arpaly. Moreover, we 
can see that there is a motivation to see if one can incorporate the concept of best 
judgment into a theory of practical rationality. If such a theory can successfully 
accommodate the case of Sam, we will have a defense of another Kantian idea – that 
rational action is behavior that is genuinely guided by the agent's understanding of the 
reasons he ought to act on. In the next chapter, I will present a new theory of action, one 
that incorporates this idea, and see if it can accommodate the case of Sam.
Chapter 4: Best Judgment and Practical Rationality

At the end of the last chapter, I motivated the idea that a proper theory of practical rationality should incorporate the concept of best judgment. The idea was that only a theory of practical rationality that incorporated the concept of best judgment could make sense of the idea that rational action is action that is guided by the agent's understanding of the reasons he ought to act on. Evidently, there is some motivation to incorporate the concept of best judgment into a theory of practical rationality. The question, however, is how exactly this new Kantian component should be incorporated. Incorporating best judgment into a theory of practical rationality is not as simple as it first might seem.

Some Background Problems

To see where some potential problems are, we can start out by looking at how such a theory of practical rationality would accommodate the case of Sam. In order for the theory to accommodate the case of Sam, it must suggest that Sam acts in accordance with his best judgment. In light of this, the theory must also suggest that Sam has a best judgment that makes his action rational, specifically a best judgment that tells him that he ought to study with friends for reasons $\psi$. However, given the way that Arpaly has illustrated the case, Sam isn't aware that he ought to study with friends for reasons $\psi$. He
is only aware that he ought to study in isolation for reasons \( \varphi \). Keeping this in mind, the solution must suggest that Sam has an \textit{unconscious} best judgment that tells him he ought to study with friends for reasons \( \psi \). The idea then would be that Sam has two best judgments. One best judgment is conscious and tells him he ought to study in isolation for reasons \( \varphi \). The other is unconscious and tells him he ought to study with friends for reasons \( \psi \). Moreover, Sam acts \textit{in accordance with his unconscious best judgment}, and hence acts rationally.

The proposal, as just outlined, might strike some as problematic for a number of different reasons. For one thing, it assumes that an agent can act in accordance with an unconscious best judgment. Someone might object to this proposal on the grounds that one cannot plausibly act in accordance with a best judgment unless it is conscious. Now, I don't think of this as being a particularly problematic assumption. Here, we should note that if we take Arpaly's case of Sam seriously, then it appears that we are committed to the idea that agents can act on \textit{reasons} (in this case, beliefs and desires) without the reasons being conscious states of the mind. Sam, after all, is allegedly an agent who acts for reasons even though the relevant beliefs and desires are not conscious states of the mind. If one simply notes that best judgments, like reasons, represent just another kind of mental state that can play a role in determining an agent's action, one shouldn't think that best judgments must be conscious in order for them to be operative in an agent's action.

Perhaps someone might object to the proposal on the grounds that it is implausible to suggest that an agent can have two best judgments \textit{at the same time}. After all, doesn't the best judgment that one ought to perform one course of action \textit{preclude} one
from having the best judgment that one ought to perform another incompatible course of action? Here, let me note that it might be impossible for an agent to consciously hold two incompatible best judgments at the same time. However, it doesn't seem impossible for an agent to hold two incompatible unconscious best judgments at the same time. Nor does it seem impossible for an agent to consciously hold one best judgment, while holding another incompatible unconscious best judgment at the same time. One should note that in Arpaly's own example, Sam is most plausibly construed as a person who consciously believes that studying in isolation will lead to success on his exams, but who nevertheless unconsciously believes that it will lead to failure. Keeping this in mind, if an agent can hold beliefs in this way, there is no good reason to think that an agent can't also hold best judgments in this way as well. Again, best judgments are just another type of mental state that can play a role in an agent's action. Given this, if we take Arpaly's case of Sam seriously, there is no good reason to deny that an agent can consciously hold a best judgment, while unconsciously holding an incompatible best judgment at the same time. In light of this, it should be noted that the aforementioned proposal is one in which Sam consciously holds the best judgment that he ought to study in isolation for reasons $\varphi$, but unconsciously holds the best judgment that he ought to study with friends for reasons $\psi$. So, the criticism doesn't seem to be a problem for the aforementioned proposal.

Perhaps someone might object to the proposal on the grounds that it makes Sam both rational and irrational in acting. Specifically, if an agent holds two incompatible best judgments, then doesn't his acting in accordance with one best

---

73 She doesn't say this explicitly, but Sam's deliberation certainly involves his belief that isolation will lead to success on exams. Moreover, Sam's unconscious reason for acting is one that is most easily construed as involving the belief that isolation won't lead him to success.
judgment also necessarily involve him acting *against* the other? In which case, aren't agents in such a situation stuck with being both rational *and* irrational in acting? This result would be problematic, since I am characterizing Sam as an agent who holds two incompatible best judgments. If this criticism were correct, I would have to conclude that Sam is both rational and irrational in his action, a conclusion that clashes with Arpaly's suggestion that Sam is rational, not also irrational. Again, I don't think that this is an insurmountable objection. Given the argument that I provided on behalf of the Kantian in the second chapter, we can see why this is this the case. There is good reason to think that Sam doesn't act *against* his conscious best judgment. Of course, if Sam doesn't act against his conscious best judgment, then we do not need to grant the claim that Sam's acting in accordance with his unconscious best judgment necessarily involves him acting against his conscious one. Given this, we can also deny the claim that Sam is both rational and irrational in acting.

There is a more concerning objection to the proposal. Someone might object to the proposal on the grounds that best judgment is the product of deliberation. Given that one assumes this claim, then one might argue that Sam cannot have a best judgment that he ought to study with friends for reasons \( \psi \), since he does not deliberate in reaching such a best judgment. Now, Sam *does* deliberate, however his deliberation yields for him the best judgment that he ought to study in isolation for reasons \( \phi \). Since Sam does not have a best judgment that he ought to study with friends for reasons \( \psi \), Sam cannot act in accordance with this best judgment either, and hence one cannot suggest that Sam acts rationally. This objection seems to be a bit more daunting. However, at this point, let me
make a suggestion. This objection can lead us to one of two possible conclusions. One possible conclusion is that Sam does not have a best judgment telling him he ought to study with friends for reasons $\psi$. According to this conclusion, the aforementioned proposal cannot work since it cannot say that Sam acts in accordance with his best judgment. The other possible conclusion is that best judgment can be the result of something other than deliberation. According to this conclusion, the assumption that grounds the initial objection is incorrect. In light of this second conclusion, the aforementioned proposal remains coherent, since it is still possible for Sam to have a best judgment telling him he ought to study with friends for reasons $\psi$.

Keeping these two options in mind, I suggest we go with the latter one. However, adopting this second conclusion places some significant restrictions on the proposed theory of practical rationality. In light of everything that has been said, the proposal would have it that Sam has two best judgments on action. One best judgment is conscious and tells him he ought to study in isolation for reasons $\phi$. The other best judgment is unconscious and tells him he ought to study with friends for reasons $\psi$. Sam acts in accordance with this unconscious best judgment, and moreover, this best judgment is one that is not reached via deliberation.74 The question we now have is this: How can a best judgment be formed, if not by deliberation. And how can an agent act in accordance with a best judgment that is unconscious? I will answer these questions next.

---

74 Here, it's important to note that I am only suggesting that the agent's best judgment is not reached via the process of deliberation. I am not suggesting that deliberation is not involved at all in the course of Sam's coming to his best judgment. This is important, because in the last chapter, I claimed that agents employ the faculty of judgment in coming to judgments about reasons for action, and that judgments about reasons for action play a necessary role in action. This is, admittedly, a kind of deliberation. However, this is separate from the kind of deliberation I am talking about currently – that is, the faculty of judgment as it is employed in the direct formation of best judgments.
How Best Judgments are Formed

How might best judgments be formed, if not by deliberation? We might be at a bit of a loss to explain how this happens. However, there are some clues that can help us construct an appropriate story. Remember that according to the proposal, Sam forms an unconscious best judgment. Keeping this in mind, I think that Arpaly's theory of practical rationality can provide us with some help. Arpaly asserts that there is such a thing as content-efficacious rationalizing cause – that mental states can cause events in virtue of their rationalizing relationship with the events. Now, what is interesting about content-efficacious rationalizing cause is that mental states can cause events without the agent's awareness being involved. Mental states can cause an event without the agent being aware of the mental states or the event in question, without the agent being aware of the rationalizing relationship between the mental states and the event, and moreover without the mental states or events being conscious states of the agent. All that is required is that the mental states be mental states of the agent, and moreover that a rationalizing relationship exist between the mental states and the event.

Of course, if mental states can cause events without the agent's awareness being involved in any way, then perhaps we have an explanation as to how an agent can form an unconscious best judgment. What would this story look like? Well, I suppose it's pretty straightforward. We start out with Arpaly's contention that mental states can cause events in virtue of their rationalizing relationship with the events. We then supplement this contention with a further one – that a rationalizing relationship exists between belief-value pairs and best judgments. Given these two contentions, we can then
suggest that *belief-value pairs cause best judgments*, in virtue of their rationalizing relationship with best judgments. Since all that is required is the *existence* of a rationalizing relationship between the agent's belief-value pairs and the agent's best judgment, we can avoid the pesky awareness requirements associated with deliberation. It is not required that the agent be aware of his *beliefs* and *values* in forming the best judgment, and it is not required that the agent be aware of the *rationalizing relationship* between his belief-value pairs and his best judgment. Moreover (and this is important), it is not required that the agent's best judgment be conscious when it is formed via content-efficacious rationalizing cause. The theory, so far, accommodates the aforementioned proposal.

Let me expand on this story by taking a few cues from the last chapter. First, let me propose that belief-value pairs exert a rationalizing causal force that corresponds to their strength as a reason for action, and moreover that the rationalizing causal force of a set of belief-value pairs that supports a certain action corresponds to the cumulative rationalizing causal force of the belief-value pairs in that set. Second, let me propose that the best judgment that is *most rationalized* by the agent's belief-value pairs is the best judgment that designates the agent's strongest reasons as the reasons he ought to act on. Third, let me propose that the set of belief-value pairs that exerts the *greatest causal force* causes a best judgment that designates those belief-value pairs as the reasons he ought to act on. Keeping this in mind, we get the result that the best judgment that is *caused* by the agent's belief-value pairs as a whole will be a best judgment that designates the strongest reasons as the reasons he ought to act on, all things being equal. Of course,
sometimes this might not happen. Sometimes non-rationalizing causes might come into play and enhance or inhibit the causal efficacy of the agent's reasons in causing best judgment. In such cases, the best judgment that the agent reaches will not be the best judgment that is most rationalized by his belief-value pairs – it will be an irrational best judgment. This is not so much of a problem, however. Sometimes agents arrive at irrational best judgments. A good theory of practical rationality should be able to explain how this can be.

There is one more detail that we need to specify before we continue. Given that we are adopting a theory of practical rationality that incorporates best judgment, I suggest we take a cue from an observation made at the end of the last chapter. In the context of Arpaly's analysis of rational action, when I speak of the agent's “best reasons,” I refer to the reasons picked out by the agent's best judgment. This is to say that under a theory of practical rationality that incorporates best judgment, “best reasons” is to be construed as a best-judgment-relative term. Keeping this in mind, an agent who acts in accordance with his best judgment will act for the best reasons. It is important to interpret best reasons in this way because a theory of practical rationality that incorporates best judgment will want to explain how best judgment accommodates the analysis of rational action given earlier. It accommodates the analysis, in part, by interpreting the agent's best reasons as being the reasons picked out by his best judgment.75

75 Admittedly, we will have to make a distinction between “best reasons”, read as the reasons that support the best action (a la Arpaly) from “best reasons”, read as the reasons picked out by the agent's best judgment. Making this distinction doesn't actually cause problems for this theory of practical rationality, but it is important to note that in accommodating the analysis of rational action that Arpaly gives, this theory of practical rationality must employ the term “best reasons” in the second sense, not the first.
What Best Judgment Causes

So far, I have presented a story illustrating how best judgments are formed. Best judgments are not solely caused by deliberation. Rather, they can also be caused by belief-value pairs, where the belief-value pairs cause the agent's best judgment in virtue of their rationalizing relationship with the best judgment. Of course, this cannot be the whole story of best judgment. After all, best judgments are supposed to play a role in explaining how agents act rationally. If we grant that best judgments play a role in explaining how agents act rationally, then it seems that best judgments must also play a causal role in the agent's acting rationally. The question is how best judgments play this causal role.

In order to figure out how best judgments play this causal role, I think it will help if we turn back to our earlier analysis of rational action. According to this analysis, rational action is action done for the best reasons and moreover done for these reasons because they are the best reasons. At the end of the last chapter, I noted that the Kantian accommodated the latter part of this analysis by suggesting that the agent's best judgment plays a role in explaining why the agent acts on his best reasons. An agent acts on his best reasons because they are the best reasons when he acts on his best reasons because his best judgment tells him they are the reasons he ought to act on. This provides us with a clue as to the causal role that best judgment plays in rational action. If best judgment plays a causal role in rational action, it plays the role of causing the agent to act on his best reasons.

There is something else that we need to remember in explaining how best
judgments cause agents to act rationally. As I noted earlier, if a best judgment plays a
causal role in the agent's acting rationally, then the best judgment plays this role in such a
way that doesn't require the best judgment to be conscious. According to the proposal,
Sam is an agent who acts in accordance with his best judgment even though the best
judgment is unconscious. This gives us another clue as to how best judgment causes the
agent to act rationally. Specifically, it suggests that Arpaly's notion of content-
efficacious rationalizing cause will be of help. If mental states can cause events purely in
virtue of the fact that a rationalizing relationship exists between them, then mental states
can cause events without the mental states being conscious states of the mind. It is
simply enough that the mental states be mental states of the agent and moreover that a
rationalizing relationship exist between the mental states and the event. That said, we can
explain how best judgments play a causal role in rational action if we employ content-
efficacious rationalizing cause.

Keeping these considerations in mind, let me lay out a story regarding the
causal role of best judgments. The story begins with the theory of practical rationality
given in the previous chapter. According to this story, belief-value pairs have
rationalizing relationships with actions. Given that mental states can cause events in
virtue of their rationalizing relationship with the events, belief-value pairs can cause the
agent to act, in virtue of their rationalizing relationship with the agent's action. So far, we
have a story which is exactly the same as the one given in the previous chapter.
However, let me add to it by suggesting the following: Although belief-value pairs and
rationalizing relationships are a necessary cause of an agent's acting, they are not a
sufficient cause. Other secondary causes are required in order to help the agent's belief-value pairs cause the agent to act. These secondary causes do not play a role by serving as the immediate cause of the agent's action. Rather, these secondary causes play the role of enhancing the causal efficacy of the agent's belief-value pairs in causing him to act.\footnote{To illustrate an example of a secondary cause enhancing the causal efficacy of a primary cause: Enzymes speed up chemical reactions that break down starch into sugars in the body. The chemical reactions might be thought of as the primary cause of the breakdown of sugars in the digestive tract. Enzymes don't directly break down sugars in the digestive tract. Rather, they speed up the chemical reactions that break down sugars in the digestive tract. Enzymes are secondary causes that enhance the causal efficacy of primary causes.}

Some of these secondary causes are of the non-rationalizing sort – addictions, obsessions and the like. One of these secondary causes is of the rationalizing sort, and this is where best judgment comes in.

How can best judgment play a secondary causal role in the agent's action? Well, in the last chapter, I proposed that belief-value judgments exert a rationalizing causal force on events, given the rationalizing relationship between the belief-value pair and the events. Keeping this in mind, we can see how best judgment might play a causal role (albeit a secondary one) in the agent's action if we propose that it plays a role as part of a belief-value pair. Let me suggest one model (though there are many models we could propose). What I would suggest here is that best judgment, in conjunction with a special value, has a rationalizing relationship with one's best reasons causing one to act. In this case, best judgment would function as a special kind of belief, a belief which pairs up with a special kind of value to cause a special kind of event – one's best reasons causing one to act. What sort of value do I have in mind here? Here, we might take a cue from Velleman and suggest something like the value of rationality – essentially a
judgment that one ought to act on one's best reasons. 
So, on this story I am constructing, best judgment and the value of rationality form a belief-value pair that has a rationalizing relationship with the agent's best reasons causing him to act. Given Arpaly's contention that mental states can cause events in virtue of their rationalizing relationship with the events, we can claim that one's best judgment and the value of rationality can cause one's best reasons to cause one to act, in virtue of their rationalizing relationship with this event.

So far then, we have a two-tiered causal story. Belief-value pairs can cause one to act, in virtue of their rationalizing relationship with the action. However, they cannot cause the agent to act by themselves, since they are not a sufficient cause of action. They need the help of secondary causes. One of these secondary causes is a necessary secondary cause, and it involves best judgment. Specifically, best-judgment and the value of rationality cause one's best reasons to cause one to act in virtue of their rationalizing relationship with this event. Let me fill in a few more details. We might suggest that action is behavior that is immediately caused by belief-value pairs, where these belief-value pairs cause the action in virtue of their rationalizing relationship with the action. Moreover, we might suggest that one acts rationally when one acts in accordance with one's best judgment – when one acts on the reasons picked out by one's best judgment. One acts irrationally when one acts against one's best judgment – more specifically, when one acts on reasons that are not picked out by one's best judgment, and

77 Alternatively, one has reason to act on one's best reasons.
Here Velleman talks about the “desire to act in accordance with reasons”. In my case, I am not proposing a desire to act in accordance with reasons, but rather a value that one has reason to act on one's best reasons – or something like this.
the reasons are contained in the domain of that best judgment.⁷⁹

Given this theory of practical rationality, we can now see how it accommodates the analysis of rational action given earlier. Rational action, according to this theory, occurs when the agent acts on the reasons picked out by his best judgment. Since the agent's best reasons are defined as the reasons picked out by the agent's best judgment, we can see why rational action is also action that is done for the best reasons. Moreover, since the agent who acts rationally acts on his best reasons, in part, because of the causal role of best judgment, we get the result that rational action is action that is done for the best reasons because they are the best reasons. The agent's best judgment and the value of rationality cause the agent's best reasons to cause him to act, in virtue of their rationalizing relationship with this event. Keeping this in mind, the account of rationality does make sense of the aforementioned analysis of rational action.

We can also see that it accommodates one of the key details that we outlined in the proposal given earlier. Given that best judgments and the value of rationality cause the agent to act on his best reasons purely in virtue of the rationalizing relationship that exists between them, a best judgment can cause an agent to act on his best reasons even when it is unconscious. Since all that is required is the existence of a rationalizing relationship, an agent's best judgment doesn't need to be conscious in order for the agent to act in accordance with it. Moreover, the theory also accommodates the Kantian requirement outlined in the previous chapter. According to the Kantian, an adequate

---

⁷⁹ One might note that this account of acting against one's best judgment differs from the account I gave in the second chapter. The account differs largely because this account doesn't incorporate the Kantian concept of choice. Without choice, the Kantian requirement that I noted earlier doesn't make much sense, since this requirement is motivated by Kant's characterization of choice as being inherently responsive to best judgment. That said, I still think the above account of acting against a best judgment is still intuitive.
theory of rationality should be able to capture the sense in which rational action is guided by the agent's understanding of the reasons he ought to act on. This new theory satisfies this requirement. Since the agent acts on his best reasons, in part, because of the causal force of the agent's best judgment, the agent's behavior is guided by the agent's understanding of the reasons he ought to act on. The agent's best judgment, after all, represents the agent's understanding of the reasons he ought to act on.

At this point, I think it would be helpful to go back and review the motivations for postulating the new theory of action and practical rationality. As I argued in the previous chapter, the Kantian claim was that in order to account for the agency in action, it was important for a theory of practical rationality to make sense of the idea that rational action is behavior that is genuinely guided by the agent's understanding of the reasons he ought to act on. This claim was motivated by the idea that a look at conscious deliberate action reveals that agents who performed such actions take themselves to be guided by their understanding of the reasons they ought to act on. Moreover, given that conscious deliberate action is plausibly the paradigm case of agency in action, there is good reason to think that conscious deliberate action ought to be the type of action that we look to in developing a theory of practical rationality. I also argued that there were significant difficulties in trying to account for the agent's understanding of the reasons he ought to act on, without appealing to best judgment. One would have to take on controversial assumptions regarding what understanding amounts to, and one would have difficulties in explaining how action is genuinely guided by this type of understanding, without appealing to best judgment. This motivated the inclusion of best judgment into a
theory of practical rationality, since best judgment was an entity that plausibly represented the agent's understanding of the reasons he ought to act on, given that best judgment had, as its content, the idea that certain reasons were the reasons that one ought to act on.

Keeping all this in mind, there are many different ways in which one could incorporate best judgment into a theory of practical rationality. This theory I have provided represents just one way in which one could incorporate best judgment. However, the main goal in providing this theory is to show that a theory of practical rationality does not, in virtue of incorporating best judgment, fail to accommodate the case of Sam. If a theory which includes best judgment is able to accommodate the case of Sam, then this shows that if there is a problem with the Kantian framework, the problem does not have to do specifically with the incorporation of best judgment, and hence the idea that rational action is action that is genuinely guided by the agent's understanding of the reasons he ought to act on. With this in mind, let us turn specifically to the case of Sam and see if our new theory of action is able to accommodate it.

Returning Back to Sam

Now that we have a working theory of practical rationality that incorporates best judgment, we can see if this new theory accommodates the case of Sam. Here is the story. Sam is an agent who has previously been in situations where he has attempted to do well on his exams by resorting to extreme isolation. This strategy has never worked
out particularly well, and Sam notices that his isolation never gets him the results that he wants, although he never reflects on this fact in any normatively significant way. Given this, when his exams come up again, Sam fails to attend to the fact that isolation has never lead to success on exams. As a result, when he deliberates, he again comes to the conclusion that he ought to study in isolation for reasons $\varphi$. Now, although Sam does come to this best judgment, Sam develops a competing *unconscious* best judgment. In particular, Sam's belief-value set (a set which includes beliefs and values that he wasn't aware of in deliberation) causes him to form a *different* best judgment, in virtue of the rationalizing relationship that exists between them, specifically a best judgment that *he ought to study with friends for reasons $\psi$*, where $\psi$ includes reasons that Sam overlooked in his prior deliberation. Moreover this unconscious best judgment serves as a secondary cause of the belief-desire pairs representing his best reasons ($\psi$) causing him to act, since they enhance the causal efficacy of his best reasons in causing him to act. As a result, Sam does not end up studying in isolation for reasons $\varphi$, *he studies with his friends for reasons $\psi$*.

Since Sam's belief-value pairs cause his best judgment *via content-efficacious rationalizing cause*, Sam can form his best judgment without his best judgment being conscious. Since Sam's best judgment and his value of rationality cause Sam's best reasons to cause him to act *via content-efficacious rationalizing cause*, Sam's best judgment can play a causal role in his action while remaining unconscious. Hence Sam can act without being aware of the reasons he ought to act on, via this best judgment. Moreover, since the belief-value pairs that represent Sam's best reasons immediately
cause Sam to act *via content-efficacious rationalizing cause*, Sam does not need to be aware of the reasons that he acts on. Since Sam acts on the reasons picked out by his unconscious best judgment, we get the result that Sam acts *rationally*. He acts rationally even though he is not aware of the reasons he ought to act on, or the reasons on which he acts. In addition to this, Sam does not act *against* his conscious best judgment either. This is because one only acts *against* one's best judgment when one acts on reasons not picked out by the best judgment *and* the reasons are contained in the domain of that best judgment. However, Sam does not act on reasons that are contained in the domain of his conscious best judgment, since he does not deliberate on such reasons when he arrives at his conscious best judgment.\(^8\) In light of this, Sam does not also act irrationally.

So it appears that the new theory of practical rationality can accommodate the case of Sam. In light of this, we have at the very least shown that *if* Arpaly's example presents a problem for the Kantian theory of practical rationality, the problem does not have to do with the idea that rational action is action that is guided by the agent's understanding of what he has most reason to do. We also have shown that the problem does not have to do with the concept of best judgment. One can incorporate these Kantian ideas into a theory of practical rationality while still accommodating the case of Sam. This second conclusion should be pretty significant. As we saw earlier, Arpaly used the case of Sam to argue that best judgment was irrelevant to a proper account of rational and irrational action, but it appears that this argument cannot get her to this conclusion after all. Of course, we aren't done with our project yet. As I see it, there are

---

\(^8\) Insert story here. Apparently, secondary causes prevented the belief-value pairs representing his strongest reasons from exerting a causal force on the creation of his best judgment. This is what gives us the license to say that they were not contained in the domain of that best judgment...
still a few criticisms that the Kantian has yet to present – criticisms that will motivate further additions to our most recent theory of practical rationality.

The First Problem – The Wallace Problem

Let me begin with a criticism that Wallace has presented for Humean theories of action – one that I think is problematic in several respects, but nevertheless instructive to examine. In “Three Conceptions of Rational Agency,” Wallace discusses what he takes to be two desiderata of theories of action. The first desideratum is what he calls the “guidance condition” which specifies that a theory of action must account for the fact that our behavior can be controlled by our understanding of what we have most reason to do. The second desideratum is that a theory of action should make sense of our subjection to rational requirements even in cases where we knowingly fail to comply with them. The theory of practical rationality I have just laid out does satisfy the guidance condition. How? Well, given that an agent's best judgment represents the agent's understanding of what he has most reason to do, and given that best judgment and the value of rationality exert a rationalizing causal force on the agent's action, it appears that an agent's behavior can in fact be controlled by his understanding of what he has most reason to do.81 Does this theory also accommodate the second desideratum? Well things do not look so good.

---

81 Incidentally, we can see that the theory that we developed in the last chapter cannot satisfy the guidance condition. Why? Well, that theory had it that belief-value pairs cause agents to act, but values do not represent the agent's understanding of what he has most reason to do, at best, they represent the agent's understanding of what he has reason to do. A value is merely a judgment that specifies that a certain consideration is a reason for action, it does not specify that the agent has most reason to do something, or that the agent ought to act on one reason instead of another. A best judgment, on the other hand, can represent the agent's understanding of what he has most reason to do, because of its special content. I should also note that there are significant difficulties (pointed out in the last chapter) in trying to accommodate Wallace's guidance condition without appealing to best judgment.
According to Wallace, many theories fail to accommodate this second desideratum because they make it the case that an agent's acting against his best judgment is dependent entirely on factors with which the agent is passive – specifically the agent's beliefs and desires. Here, we should be careful to distinguish passivity, as used in previous chapters, from passivity as used in Wallace's criticism. In the previous chapters, Kantians spoke of passivity as referring merely to the involvement of the agent in action – agents were considered passive with respect to their beliefs and desires because they weren't considered to be involved (in whatever way) with respect to the existence of their beliefs and desires. Here, Wallace speaks of passivity as referring to the ability of the agent to control the causal factors that influence action. When speaking of beliefs and desires as passive, Wallace suggests that such states are not “under our direct control” in the sense that we cannot control their causal influence in action.82 Keeping Wallace's notion of passivity in mind, if an agent acts against his best judgment, and his acting against his best judgment is entirely the result of factors with which he is passive, then it appears that there is no way that the agent could have acted in accordance with his best judgment. Since there is no way the agent could have acted in accordance with his best judgment, it appears that rational requirements (i.e., normative requirements) do not apply to him. These rational requirements demand that the agent ought to act in accordance with his best judgment, but there can be no “ought” without a “can.” In light of this, if an agent is entirely passive with regards to his acting against his best judgment, there is an obvious sense in which he could not act in accordance with his best judgment. So, rational requirements do not apply to agents who act against their best judgments.

82 TC, 224.
Wallace suggests that these sorts of views fail to make sense of the idea that “rational requirements can retain their normative force for agents who deliberately flout them.” However, we should note that the actual problem that Wallace has should be that these views fail to make sense of the idea that rational requirements retain their normative force for agents, period. After all, if an agent acts in accordance with his best judgment, and his acting in accordance with his best judgment is entirely the result of factors with which he is passive, then it appears that there was no way the agent could have acted against his best judgment either. In such a case, it appears that rational requirements (i.e., normative requirements) do not apply to him either. This is a point that is similar to one that Kant made in the *Groundwork*. Angels are constructed such that they are destined to choose only that which their best judgment tells them to, and so Kant concludes that imperatives (i.e., rational requirements that signify an ought) do not apply to them. Keeping all this in mind, we can see that the real problem just is that rational requirements cannot retain their normative force for agents in general, in the case that their actions are determined by factors with respect to which they are passive.

In light of this, we can see why the theory I have just outlined might fail to accommodate this second requirement given by Wallace. After all, whether the agent acts in accordance with his best judgment or against it is dependent on two types of causal factors – the causal strength of his belief-value pairs, and the causal strength of his best judgment and the value of rationality. The causal strength of the agent's belief-value pairs is determined solely by the strength of those belief-value pairs as reasons for action. Since the agent is passive with regards to how strong his belief-value pairs are as

83 TC, 236.
reasons for action, it is evident that the agent is passive with regards to the causal strength of his belief-value pairs. After all, the strength of his belief-value pairs as reasons for action is entirely a function of an *independently existing rationalizing relationship* that exists between the belief-value pairs and the action in question. Moreover, since the agent's best judgment and the value of rationality just are another belief-value pair, it seems we can only reasonably conclude that the agent is likewise passive with regards to the causal strength of this causal factor as well. Keeping this in mind, it's not clear that this theory of practical rationality can meet this second desideratum.

What is Wallace's solution to these problems? Well, he suggests that “there is an important class of motivational states that are directly subject to our immediate control. Familiar examples from this class of motivations are such phenomena as decision and choice. Ordinarily we think of decisions and choices not merely as states to which we happen to be subject. Rather they are states for which we are ourselves directly responsible, primitive examples of the phenomenon of agency itself.”\(^{84}\) Wallace's solution is essentially that we need to bring *choice* back into the picture. If there is a motivational state that the agent is active with regards to, and such a state plays a role in determining whether an agent acts in accordance with or against his best judgment, then we have an explanation for how agents can be subject to normative requirements. Choice is the motivational state that he sees as playing this role.

If we take Wallace's criticism and solution to be plausible, then it appears that even our most recent theory of practical rationality is problematic. This theory of practical rationality has a problem in that it takes one's acting in accordance with or

---

84 TC, 236-237
against one's best judgment to be entirely the result of causal factors with respect to
which the agent is passive. Given this, it cannot make sense of the idea that agents are
subject to normative requirements. Keeping this in mind, in order to avoid Wallace's
criticism, we have to replace these causal factors with something like choice – a primitive
motivational state with respect to which the agent is active. A motivational state that can
either move the agent to act in accordance with his best judgment or against it, and one
which moreover makes sense of the claim that we, as agents, are susceptible to normative
requirements.

A Re-Examination of Wallace, and a New Motivation for Incorporating Choice

Now, while I do think that Wallace's criticism is interesting, I think that it
suffers a problem in that it employs the active/passive distinction in arguing for the
incorporation of choice. The problem is similar to a problem that I pointed to in the last
chapter. Wallace does not actually provide an independent standard showing which
states agents are active with respect to, and which states agents are passive with respect
to. If this is right, then it's not clear that he can legitimately use the active/passive
distinction in making a substantive criticism of Arpaly's theory, or any other theory for
that matter. After all, if one cannot actually distinguish between states based on activity
and passivity, without stipulation, then it's not clear that the distinction is a real one (after
all, what are distinctions for, but to help distinguish between things?). If this is right,
then it's not clear that one can use the active/passive distinction, even in its current
incarnation, to launch a criticism against any theory of practical rationality.
At this point, I think it is best to scrap Wallace's argument entirely and look at a much more simple argument for the incorporation of choice – one which I think stays more true to actual Kantian motivations for incorporating choice. Here, it's important to go back and look at what was so special about choice in the first place. As the Kantian sees it, choice is distinct in that it (1) has as its content a principle stating the action and reasons that the agent commits himself to act on, and (2) is necessarily accompanied by the “I think” of self-consciousness – in other words, the agent is necessarily aware of the reasons that he commits himself to act on. In light of this characterization of choice, what is the problem with our latest theory of practical rationality? Well, the problem with our latest theory of practical rationality (and with any theory that doesn't include choice) is that it fails to secure another sense in which action involves agency. It fails to secure the sense in which action proper requires that action be genuinely guided by the agent's understanding of the reasons he commits himself to act on. (Here, we should be careful to distinguish the agent's understanding of the reasons he commits himself to act on, from the agent's understanding of the reasons he ought to act on, and the agent's understanding of the reasons he has for action). Such an understanding, suggests the Kantian, is necessary in order for the actions of the agent to truly count as representing the agent. Incidentally, we can see why a theory that incorporates choice can secure this understanding. Choice is a mental state that has as its content the idea that one commits oneself to performing an action for a reason, so it counts as representing an understanding of what reasons one commits oneself to act on. It is a state which is necessarily accompanied by the “I think” of self-consciousness, so it counts as
representing the agent's understanding of the reasons he commits himself to act on. Lastly, it is a state that plays a causal role in action, and so counts as a state that genuinely guides the agent's action.\textsuperscript{85}

Why is it that one would take it to be the case that action is behavior that is genuinely guided by the agent's understanding of the reasons he commits himself to act on? As before, I think that this Kantian claim is motivated by two other claims: that agents engaged in conscious deliberate action take their action to be guided by this understanding, and that conscious deliberate action is the paradigm case of action. Regarding the first claim, the idea is that when we look at examples of conscious deliberate action, agents who perform such actions take their actions to be guided by their understanding of the reasons they commit themselves to act on. Moreover, such agents cannot deliberately act until they come to such an understanding. Regarding the second claim, the idea is that conscious deliberate action ought to be considered the paradigm case of action because agency is taken to be a definitive characteristic of action, and actions are taken to be most expressive of agency when they are conscious deliberate actions.

Now, if one takes this argument to be plausible, we have some motivation to incorporate choice back into a theory of practical rationality. Of course, incorporating

\textsuperscript{85} One might wonder how this criticism relates to the view that I outlined in the previous chapter. Here, I should note that the criticism does not contradict the criticisms given in the previous chapter. Each criticism provides an argument that a different type of understanding ought to be addressed by a theory of practical rationality, where these understandings are not mutually exclusive. However, the criticism given in this chapter does show that the theory of practical rationality given in the last chapter is faulty, in that it isn't able to account for the agent's understanding of the reasons that he commits himself to act on. As I see it, any theory that doesn't include choice will have significant difficulties in accounting for such an understanding. As my reasons for thinking this are similar to the reasons that I pointed out for Arpaly's theory in accommodating the other types of understanding, I will not repeat them.
choice back into the picture is not an easy task. As just noted, when Kant speaks of choice, he thinks of the agent's awareness as necessarily being involved. The connection between choice and consciousness is the very thing that makes it the case that an agent's choice represents the agent's understanding of the reasons that he commits himself to act on. This is problematic, because Kant would have it that rational action is behavior that is caused by the agent's choosing to act on the reasons picked out by his best judgment. In light of this, Kant thinks that agents who act rationally are necessarily aware of the reasons that they commit themselves to act on. However, Sam is an agent who allegedly acts rationally, yet is unaware of the reasons he commits himself to act on. In light of this, if Sam chooses to study with friends for reasons ψ, then his choice is unconscious. This puts us in a bit of a dilemma. In order to accommodate the Kantian, choices must be conscious and hence accompanied by an awareness of the reasons that one commits oneself to act on. In order to accommodate the case of Sam, choices must be unconscious and hence unaccompanied by an awareness of the reasons that one commits oneself to act on. In order to get out of this dilemma, I will have to show that unconscious choices of the agent can represent the agent's understanding of the reasons he commits himself to act on. This is not an easy task, but I think a modified version of the original Kantian theory of practical rationality can get us there.

The Second Problem – Best Judgment and Understanding

Earlier in this chapter, I met the Kantian half way by incorporating best judgment into the theory of practical rationality. However, I accomplished this in such a
way that eliminated deliberation as a necessary cause of best judgment. Moreover, I incorporated best judgment in such a way that did not require that best judgments have a necessary connection to the agent's awareness – i.e., that best judgments be conscious. Keeping this in mind, there is a significant problem that Kantians are going to have with this theory of practical rationality.

I earlier claimed that a good theory of practical rationality should accommodate the idea that rational action is guided by the agent's understanding of the reasons he ought to act on. However, there is a sense in which best judgment, as characterized by this most recent theory, does not actually accommodate this idea. According to the Kantian, an understanding of the reasons one ought to act on does not count as the agent's understanding of the reasons he ought to act on unless such an understanding is connected to the agent's self-consciousness. Here, Kantians will point to a statement made by Kant, one that I invoked in the first chapter:

“The 'I think' must accompany all my representations, for otherwise something would be represented in me which could not be thought; in other words, the representation would either be impossible, or at least be, in relation to me, nothing...”\(^{86}\) (my emphasis)

Here, Kant explicitly suggests that in order for a representation to be, in relation to the agent something, it must be accompanied by the “I think” of self-consciousness. In light of this claim, the Kantian will suggest that one cannot really claim that rational action is guided by the agent's understanding of the reasons he ought to act on, unless best judgment is necessarily connected to consciousness. Given this, it appears that our most recent theory of practical rationality fails. It does not properly

\(^{86}\) CP, B:131.
accommodate the notion that rational actions are genuinely guided by the agent's understanding of the reasons he ought to act on. Of course, one could respond to this criticism by creating a theory of practical rationality that incorporated conscious best judgment, but this leads to another problem. Sam is an agent who is not aware that he ought to act on reasons ψ, so if he has a best judgment at all, it is an unconscious best judgment. This puts us in a bit of a dilemma. In order to accommodate the Kantian, best judgments must be conscious and hence accompanied by an awareness of the reasons that one ought to act on. In order to accommodate the case of Sam, best judgments must be unconscious and hence unaccompanied by an awareness of the reasons that one ought to act on. In order to get out of this dilemma, I will have to show that unconscious best judgments of the agent can represent the agent's understanding of the reasons that he ought to act on. This is not an easy task, but again, I think a modified version of the original Kantian theory of practical rationality can get us there.

**Preview of the Next Chapter**

Given what has been said, the task I have is difficult. If we adopt choice and best judgment wholesale from the Kantian theory, then we fail to properly accommodate the case of Sam. So, if we do bring choice and best judgment back into the picture, some modification of these concepts will be required. I will have to show that choice and best judgment can be unconscious, and yet still be representative of the agent's understanding of the reasons he commits himself to act on, and representative of the agent's understanding of the reasons he ought to act on, respectively.
In the next chapter, I will show how this task can be accomplished, by presenting a final theory of practical rationality. Given that this final theory of practical rationality accommodates Arpaly's case of Sam and adequately accommodates these latest Kantian sentiments, we will have a defense of a Kantian theory of practical rationality outlined in the first chapter. Of course, given that this final theory does a better job in accommodating the case of Sam, we will have a Kantian theory of practical rationality that counts as an *improvement* over the original Kantian theory of practical rationality, as presented in the first chapter.
Chapter 5: The Kantian Theory of Practical Rationality, Revised

In the previous two chapters, I set out to defend certain aspects of the Kantian theory of practical rationality. In the third chapter, I defended the idea that action is behavior that is guided by the agent's understanding of what he has reason to do, and specifically the idea that judgments about reasons for action ought to be incorporated into a theory of practical rationality. In the fourth chapter, I defended the idea that rational action is action that is guided by the agent's understanding of the reasons he ought to act on, and specifically the idea that best judgment ought to be incorporated into a theory of practical rationality. In both cases, I outlined the Kantian motivations for incorporating these ideas into a theory of practical rationality. I then constructed a theory of practical rationality that incorporated these ideas. I tested these theories against the case of Sam, and since they succeeded in accommodating the case of Sam, I concluded that the Kantian theory of practical rationality did not fail to accommodate the case of Sam in virtue of its incorporating these ideas. At the end of the last chapter, I outlined two problems for the latest theory of practical rationality. The theory did not accommodate the idea that action is behavior that is guided by the agent's understanding of the reasons that he commits himself to act on, and it did not *properly* accommodate the idea that action is behavior that is guided by the agent's understanding of the reasons he ought to act on.
These problems created a bit of a dilemma. According to the Kantian, the solution to these problems is had by incorporating best judgment and choice in such a way that requires that they be conscious. This is to say that the Kantian theory requires that agents who form best judgments and choose be aware of the reasons that they ought to act on, and be aware of the reasons that they commit themselves to act on. After all, suggests the Kantian, one cannot show how action is guided by the agent's understandings unless one connects the representations that guide the agent's action to the agent's consciousness. This is problematic, because it appears that in order for one to accommodate the case of Sam, one must suggest that Sam forms an unconscious choice and an unconscious best judgment. This dilemma led us to a significant task: We must show that best judgments and choices can be unconscious, but nevertheless representative of the agent's understanding of the reasons he ought to act on, and the reasons he commits himself to act on, respectively.

In order to accomplish this task, I want to go back to the Kantian theory of practical rationality outlined in the first chapter and offer up a modified version of it. The modified theory is a lot like the original Kantian theory, except that it offers up a story regarding how best judgments and choices can be formed unconsciously. Given that best judgments and choices can be formed unconsciously, the modified Kantian theory of practical rationality successfully accommodates the case of Sam. In light of this, we have a defense of a kind of Kantian theory of practical rationality. Of course, given that such a theory has it that best judgments and choices can be formed unconsciously, we will run afoul of the Kantian claim that mental states must be conscious in order to be
representative of the agent's understanding in action. So, the theory that I present does represent a departure from the traditional Kantian account. However, as I see it, the departure is in fact motivated by concerns that are separate to the case of Sam. Moreover giving up on this claim helps us to preserve more significant Kantian commitments – that action is behavior that is genuinely guided by representations that reflect the agent's understanding of what he is doing, and that rational action is action that accords with one's best judgment and irrational action is action that goes against one's best judgment.\(^7\) So long as the theory preserves these central Kantian commitments, the modified theory represents a defense of a kind of Kantian theory. Moreover, given that this theory has a greater explanatory power than the original Kantian theory, it will be an improvement on the original Kantian theory of practical rationality.

The Kantian Theory of Practical Rationality, and the Original Problem

To start out, let's return back to the Kantian account of how best judgments are formed. According to the Kantian theory of practical rationality, best judgments are formed via deliberation. In reaching a best judgment, Kant suggests that an agent begins with awareness of considerations. Given one's awareness of such considerations, judgment and the imperatives come into play to make the agent aware of reasons for action. Given one's awareness of reasons for action, judgment and the imperatives come into play again to make the agent aware of the reasons he ought to act on. This view outlined by Kant seems to suggest that whenever an agent acts, he is minimally aware of

\(^7\) Again - acting in accordance with one's best judgment is to be taken as a necessary condition of rational action, whereas acting against one's best judgment is to be taken as a sufficient condition of irrational action.
considerations, reasons for action and the reasons he ought to act on. This is problematic, however, because agents don't always seem to be aware of these things when they act.

Herman addressed this problem by supplementing Kant's theory of practical rationality with the agent's conception of the good. The judgments that an agent forms through deliberation are retained by the agent after deliberation has completed, and these judgments form the agent's conception of the good. According to Herman, such judgments can interact with the agent's perception of his situation and principles of salience to make the agent immediately aware of reasons for action and their comparative weight. Given one's awareness of reasons for action and their comparative weight, judgment and the imperatives come into play to make the agent aware of the reasons he ought to act on, via best judgment. Given that the agent's conception of the good can operate in this way, the need for deliberation in action is greatly reduced. Of course, as Herman sees it, deliberation cannot be eliminated entirely. In coming to a best judgment, the agent is still minimally required to be aware of reasons for action, their comparative weight, and he ends up being aware of the reasons he ought to act on.

Of course, even these minimal awareness requirements create problems for the Kantian theory of practical rationality. In order for the Kantian theory to accommodate the case of Sam, it must claim that Sam acts in accordance with his best judgment, since Sam acts rationally. However, in order to claim this, the Kantian must claim that Sam has this best judgment in the first place – namely a best judgment that tells him he ought to study with friends for reasons ψ. However, Sam doesn't seem to be
aware that he ought to study with friends for reasons ψ, and given how Arpaly illustrates the case, *Sam has never deliberated in coming to this best judgment at any time*. The awareness requirements that Kantians associate with deliberation are not ones that Sam satisfies in the course of his performing his action. In light of this, the Kantian cannot say that Sam has a best judgment that he ought to study with friends for reasons ψ. However, if this is the case then the Kantian cannot say that Sam acted in accordance with his best judgment, and likewise cannot say that he acted rationally. It seems then that the Kantian account cannot accommodate the case of Sam without some kind of modification.

**Thesis 1 – Best Judgments and the Agent's Conception of the Good**

What is the Kantian to do? Evidently the Kantian theory of practical rationality needs to be modified in some way. At this point, I want to return back to Herman's notion of the agent's conception of the good and argue for two modifications that I think will be helpful. One concerns the *kinds* of judgments that are contained in the agent's conception of the good, and the other concerns how *changes* are made to the agent's conception of the good.

As noted earlier, the agent's conception of the good essentially consists of the set of value judgments that the agent forms through prior instances of deliberation. It contains the agent's value judgments about reasons for action, and moreover value judgments about the comparative weight of such reasons. Now, what I would like to argue for first is that *best judgments* are also contained in the agent's conception of the
good. Of course, Herman seems to suggest that best judgments aren't contained in the agent's conception of the good. This is evident because the agent's conception of the good, under her view, doesn't directly bring the agent to a best judgment on action. At most, it directly brings the agent to awareness of reasons for action and their comparative weight. Best judgment only comes into the picture after the agent employs the faculty of judgment and the imperatives, given his awareness of reasons for action and their comparative weight.

However, it should be noted that Herman doesn't actually argue that best judgments can't be contained in the agent's conception of the good. As I see it, there are very good reasons to think that best judgments are contained in the agent's conception of the good. Herman sees the agent's conception of the good as consisting of normative judgments that are the product of the agent's deliberation. However, best judgments just are normative judgments that are the product of the agent's deliberation! They are specifically normative judgments that point out how one ought to act, all things considered.  

Perhaps, one might think that best judgments can't be contained in the agent's conception of the good specifically because best judgments are time-indexed – that is, they say what reasons an agent ought to act on, now. Given that the agent's conception of the good contains judgments that can be employed on multiple occasions, one might think it impossible then for best judgments to be contained in the agent's conception of the good. Regarding this, it's not clear to me that a Kantian would agree with the

88 When I say “all things considered” I mean all things the agent has considered in forming the best judgment.
assumption. For the Kantian, it is not part of the *content* of the agent's best judgment that he ought to act on certain reasons *now*. Rather, when the agent consciously holds a best judgment, he is *now* aware that he ought to act on certain reasons. In light of this, best judgments really aren't time-indexed in the way described earlier. They are general, and so there is no reason to think that best judgments can't also be retained in the agent's conception of the good as well.89

**Thesis 2 – Changes to the Agent's Conception of the Good**

So far I've claimed that best judgments are contained in the agent's conception of the good. I don't think that this claim is very controversial, in part because Herman's own view of the conception of the good leads us naturally to this claim. Let me now make a more controversial claim. Specifically, I would like to claim that *changes* to the agent's conception of the good can be made in ways that don't *directly* require deliberation. Why might this claim be controversial? Well, if we look at the original Kantian theory of practical rationality given earlier, it appears that deliberation is the only force which shapes the agent's conception of the good. It is through deliberation that one forms value judgments, and it is these value judgments that constitute the agent's conception of the good. I earlier modified how we are to think of the agent's conception of the good by suggesting that best judgments are also contained in it. However, even in this case, it appears that best judgments can only enter the agent's conception of the good

---

89 Of course, in order for the best judgment to be effective *in a particular action* it must be the case that the agent *applies* the best judgment to a particular case so that he can become aware that he *now* ought to act on certain reasons. However, we can, presumably, make use of the rest of Herman's framework in laying out how this is possible – specifically using her concepts of “perception of the agent's situation” and “principles of salience”.

136
through deliberation.

This picture is certainly compelling in some sense, but I think it may lead one to an overly simplistic view regarding how the agent's conception of the good is shaped. Specifically, it might lead one to the view that all changes to the agent's conception of the good are ones that directly require deliberation, and I would like to argue that this is incorrect. In order to see why it might be incorrect, let me present some examples that I think challenge this view:

(5) Suppose that Sally values drinking a certain fruit juice, because she believes it has a variety of cancer-fighting benefits. As a result of this, Sally also values a variety of other things as well – for instance she values finding out where she can get this juice, and she values keeping a good stockpile of this juice. Unfortunately, Sally reads an article and finds out that a major study has concluded the fruit juice to be a cancer-causer instead. Sally notices this, and through conscious deliberation, she ceases to value drinking the fruit juice. However, without her deliberating on it, Sally ceases to value finding out where she can get it, and ceases to value keeping stockpiles of it as well. It wasn't that Sally was consciously aware of these values, and aware of their relationship with the value of drinking the fruit juice, and through deliberation ceased to value them. She just stopped valuing them.

(6) Suppose that Gary values eating both kale, spinach and mustard greens in virtue of their antioxidant properties – and moreover he values mustard greens more than he values spinach, and values spinach more than he values kale, in virtue of the fact that mustard greens have more antioxidant properties than spinach, and spinach has more antioxidant properties than kale. Let us also suggest that of the three, Gary values mustard greens the most of the three vegetables because of its antioxidant properties. Unfortunately, Gary reads an article and finds that the antioxidants in mustard greens are not readily absorbed by the body, and so the antioxidant benefits of eating mustard greens are next to negligible. Upon seeing this, he deliberates and ceases to value eating mustard greens. However, without his deliberating on it, Gary comes to value spinach the most. It wasn't that he consciously deliberated about the various antioxidant benefits of mustard greens, spinach
and kale, and came to a new judgment that spinach was now the most valuable of the three – *he just came to value spinach the most.*

What I would like to suggest with these examples is that under any plausible construal of the agent's conception of the good, not all changes to the agent's conception of the good are directly the result of deliberation. In the first case, Sally rejected one of her “upstream” values via deliberation, and as a result some of her “downstream” values were *unconsciously* rejected.\(^90\) In the second case, Gary rejected one of his “upstream” values via deliberation, and as a result, some of the “downstream” comparative value judgments *changed* in response to his rejection of his “upstream” value. Here, the terms “upstream” and “downstream” refer to the justificatory relation that the two values have to one another. An “upstream” value is the *justifier* of a “downstream” value. The “stream” refers to the justificatory relationship between the two values.

Keeping this in mind, if we take these examples to be plausible, we should reject the notion that all changes to the agent's conception of the good directly require deliberation. Notice that when I suggest this, I do not thereby endorse the proposal that *none* of the changes to the agent's conception of the good directly require deliberation. I am only suggesting that some of the changes to the agent's conception of the good can take place without deliberation being directly involved. Moreover, let me emphasize that I do not mean to suggest that deliberation isn't involved at all in these changes. I am only suggesting that deliberation isn't *directly* involved.

Of course, at this point, some will ask for a more explicit account of what sorts of changes directly require deliberation and what sorts of changes don't directly

---

require deliberation. I don't think I will be able to lay out an exhaustive account of this, however, let me suggest a general rule that one might find attractive. As I see it, there are three characteristics that are common to the examples that I offered up earlier. First, the unconscious changes to the agent's conception of the good are byproducts of changes that the agent makes via deliberation. Sally unconsciously ceases to value finding a supplier of the fruit juice only after she ceases to value drinking the fruit juice via deliberation. Gary unconsciously comes to value spinach the most only after he ceases to value eating mustard greens via deliberation. Second, the unconscious changes are all downstream changes that maintain consistency with an upstream change that the agent makes via deliberation. Sally's value of finding a supplier of the juice is justified by her value of drinking the juice. In light of this, Sally's ceasing to value finding a supplier of the juice maintains consistency in her conception of the good, given that she ceases to value drinking the juice. Gary's valuing mustard greens the most is justified in part by his value of eating mustard greens. In light of this, Gary's valuing spinach the most maintains consistency in his conception of the good, given that he ceases to value eating mustard greens. Third, the unconscious changes are not additive in that they do not add new value judgments to the agent's conception of the good. Rather, they eliminate components in the agent's conception of the good, or they modify existing judgments in order to maintain a consistency in the agent's conception of the good.

Keeping this in mind, we might think that the following pair of rules provides us with a way to determine which changes directly require deliberation and which ones don't: (1) Changes to the agent's conception of the good that require the addition of new
value judgments require deliberation. (2) Changes to the agent's conception of the good can occur without deliberation when they are downstream byproducts of changes that the agent makes via deliberation. However, these changes cannot be additive and only serve to maintain “downstream consistency” with upstream changes that are made.

Now, if we take the aforementioned examples to be psychologically plausible, then it appears that changes to the agent's conception of the good can occur without deliberation being directly involved. Moreover, if we take these new rules to be consistent with the aforementioned examples, then it appears we have a way to distinguish what sorts of changes to the agent's conception of the good directly require deliberation and what sorts of changes don't directly require deliberation. We now have the beginnings of a new theory of practical rationality that allows changes to the agent's conception of the good to occur without deliberation being directly involved. However, one might wonder if this moves us into extremely un-Kantian territory. After all, I have now partially severed the connection between deliberation and the agent's conception of the good. In the next section, I will explain why my proposal here is not un-Kantian at all.

**Herman and the Function of Deliberation**

In order to see why my claim in the previous section is not un-Kantian, I think it will help to look at the role that Herman sees deliberation playing in her theory. In “Making Room for Character,” Herman explicitly discusses why she thinks

---


140
deliberation is important to a theory of action in the first place. She notes:

“What gets lost in the strategic maneuvering at the level of theory is the hard work of moral deliberation that is central to a moral life: the engagement with multiple moral considerations present in an agent's current or anticipated circumstances of action...However well one is brought up, however complete the internalization of a regulative moral motive, what one will have to know, or how one may have to rethink one's own values, cannot be predicted. There is thus reason internal to the moral phenomena to want a characterization of moral life that supports a certain degree of flexibility in both judgment and motivation...What raises a question about the flexibility of this arrangement of judgment and motivation is the apparent absence at the limit of a way (for the agent) to criticize the sensitivity itself – for it to take itself as the object of its own critical regard.”

Here, Herman criticizes a virtue theory of action in which deliberation is absent and action proceeds directly from the agent's values (something like the agent's conception of the good). Her suggestion is that without deliberation, there is no way for the agent to deal with cases that challenge his conception of the good. As she sees it, there will always be cases in which the agent's conception of the good does not have the resources to tell the agent what he ought to do. Regardless of how rich one's conception of the good is, there will be circumstances in which it needs to be changed, and a certain “flexibility” is required in order to deal with such circumstances. There will be a need to “criticize the sensitivity itself” and for the conception of the good to “take itself as the object of its own critical regard.” Deliberation, suggests Herman, is the very process that takes the agent's conception of the good as its object, and changes it to help the agent come to grips with cases of practical perplexity.

What sorts of situations bring about the need for deliberation? Herman doesn't give us a list of the types of situations that require deliberation in this passage. However, in other passages, Herman does offer up some examples. For instance, she

92 MR, 54.
thinks that deliberation is especially necessary in cases where the agent is aware of a variety of reasons for action but does not know which reasons take priority. Herman also suggests in several places that deliberation is necessary in cases where the agent is confronted with new moral facts – cases in which previously unknown considerations present themselves to the agent as requiring action. These are cases in which the agent's conception of the good does not have the resources to tell the agent how he ought to act. In such cases, new judgments are necessary and deliberation is the force that supplies the agent with the required new judgments. In cases where the agent doesn't know which reasons take priority, he must deliberate and come to the new judgment that some reasons take priority over others. In cases where the agent hasn't considered a new moral fact, he must deliberate and come to a new judgment that it constitutes a reason for action. As I read Herman, deliberation is necessary in adding new judgments to the agent's conception of the good.

However, keeping this in mind, we can see that my earlier claim about unconscious changes to the agent's conception of the good is consistent with her view of the role of deliberation in shaping the agent's conception of the good. She seems to suggest that deliberation is required in order to add new judgments to the agent's conception of the good. I suggest that changes to the agent's conception of the good can happen without deliberation being directly involved, but only in cases where they are not additive and moreover only eliminate or modify existing judgments in order to maintain overall consistency in the agent's conception of the good. One can adopt my previous

93 PMJ, 146.
claim while still retaining Herman's view regarding the role of deliberation in shaping the agent's conception of the good. In both of the examples that I previously presented, there is a strong sense in which the agent's conception of the good already has the resources to “criticize itself and take itself as the object of its own regard.” In the first case, Sally's conception of the good already contains in it the understanding that her “downstream” value is rationally supported by her “upstream” value. The “downstream” value is thus eliminated in light of this understanding, and in light of the fact that her “upstream” value has been eliminated. In the second case, Gary's conception of the good already contains in it the understanding that his “downstream” comparative value judgment is rationally supported by his “upstream” value. The “downstream” comparative value judgment is thus modified in light of this understanding, and in light of the fact that his “upstream” value has been eliminated.

**Bringing Back Arpaly's Content-Efficacious Rationalizing Cause**

So far, I've claimed that some changes to the agent's conception of the good don't directly require deliberation. I've also argued that this claim is consistent with the role that Herman sees deliberation playing in a theory of action. At this point, one might wonder exactly how it is that these unconscious changes come to take place – that is, what is it that directly causes the changes, and how it is that deliberation indirectly plays a role in the changes. Here, I would like to propose a story in which Arpaly's content-efficacious rationalizing cause plays a role in explaining how this is possible.

Let me start out by going back to the case of Sally. As I suggested, Sally
unconsciously ceases to value finding a supplier of the fruit juice once she ceases to value drinking the fruit juice. How is it that Sally ends up ceasing to value finding a supplier of the fruit juice? What I would like to suggest is that Sally's ceasing to value finding a supplier of the fruit juice is caused in virtue of a rationalizing relationship that exists between her value of drinking the fruit juice, her belief that finding a supplier of it will aid her in drinking it, and her value of finding a supplier of it. Specifically, it is the existence of this rationalizing relationship that makes it the case that when she ceases to value drinking the fruit juice, she thus also ceases to value finding a supplier of it. This is to say that content-efficacious rationalizing cause does exist. However, rationalizing relationships specifically exist between “upstream” belief-value pairs and “downstream” values contained in the agent's conception of the good. Moreover, content-efficacious rationalizing cause cannot, on my view, result in the creation of new values in the agent's conception of the Good. It can only result in the elimination of values or the modification of values contained in the agent's conception of the Good, where such modifications maintain consistency with “upstream” belief-value pairs.

Let me add to this story about content-efficacious rationalizing cause. Although the existence of a rationalizing relationship is a necessary condition for the causing of unconscious changes to the agent's conception of the good, it is not a sufficient condition. It is also necessary that the agent previously deliberate from the relevant “upstream” belief-value pair to the relevant “downstream” value. Let me illustrate by looking back at the case of Sally. Sally's value of drinking the juice and her belief that finding a supplier of it will aid her in drinking it, have a rationalizing relationship with
her value of finding a supplier of the juice. Now, although a rationalizing relationship exists between these states, the mere existence of this rationalizing relationship cannot cause Sally's ceasing to value finding a supplier of the fruit juice when she ceases to value drinking it, at least not by itself. The rationalizing relationship can only cause her ceasing to value finding a supplier of the fruit juice when it is the case that she has previously deliberated from her value of drinking the fruit juice and her belief that finding a supplier of it will aid her in drinking it, to her value of finding a supplier of it.\(^5\)

Here, the agent's deliberation does not play a direct causal role in Sally's ceasing to value finding a supplier of the fruit juice. Nevertheless, her deliberation does play a necessary indirect causal role in that it allows the existence of a rationalizing relationship to make it the case that Sally ceases to value finding a supplier of the fruit juice.

Note that content-efficacious rationalizing cause, on the view outlined, can only be effective in the case that the agent is rational in his deliberation. If the agent rationally deliberates from an upstream belief-value pair to a downstream value, a rationalizing relationship will exist between the upstream belief-value pair and the downstream value. This makes it the case that content-efficacious rationalizing cause can be effective in eliminating the downstream value when one of the upstream mental states (either the value or the belief) is eliminated, since a rationalizing relationship does exist between these states. However, if the agent deliberates irrationally from a belief-value pair to a new value, the new value itself cannot be eliminated via content-efficacious cause.

---

\(^5\) One might be tempted to think that the existence of a rationalizing relationship between a pair of values shows that the agent has deliberated from the “upstream” value to the “downstream” value, but this is not always the case. After all, it is possible that the agent arrived at the “downstream” value through irrational deliberation, or that the agent arrived at the “downstream” value rationally, but via a different belief-value pair.
rationalizing cause, as a rationalizing relationship does not exist between the belief-value pair and the new value. That said, this isn't such an implausible result. Kantians take mistakes in judgment to be partly the result of non-rationalizing external causes. In light of this, it is not a surprise that the elimination of a member of the belief-value pair does not result in the elimination of the new values, when the agent reaches the downstream value in an irrational manner.

Let me try and explain the overall story that I have in mind from a slightly different perspective. According to Herman, the agent's conception of the good is a set of value judgments that is structured by deliberation. This is to say that deliberation (1) provides the agent's conception of the good with the value judgments contained in it, and (2) establishes relationships (both priority relationships and deductive relationships) between the value judgments contained in it. I accept Herman's account of the agent's conception of the good, but I go further by supplementing it with a new claim: Through rational deliberation, the agent's deliberation also instills in the agent's conception of the good a special kind of sensitivity that operates via content-efficacious rationalizing cause – a sensitivity that itself can have a role in shaping how the agent's conception of the good is structured. The three rules for when unconscious changes to the agent's conception of the good can take place, outlined earlier, are rules that characterize the nature of this sensitivity and how it can shape the agent's conception of the good.

Keeping my story about unconscious changes to the agent's conception of the good in mind, it should be clear now why I earlier asserted that some changes to the agent's conception of the good indirectly involve deliberation. As we can see here, there
is a significant sense in which these sorts of changes do require deliberation after all. Content-efficacious rationalizing cause can only be effective in the case that the agent deliberates from the relevant upstream belief-value pair to the relevant downstream value. Moreover, rationalizing relationships will, for the most part, only exist between upstream belief-value pairs and downstream values contained in the agent's conception of the good when the agent rationally deliberates from the upstream belief-value pairs to the downstream values. That said, we can see that according to this new story, the unconscious changes that take place in the agent's conception of the good still indirectly require deliberation.

The Importance of What has been Said

So far, I have argued for two main claims. The first is that best judgments are components of the agent's conception of the good. The second is that changes to the agent's conception of the good can be take place without deliberation being directly involved. Of course, neither of these claims will be helpful unless they can help the Kantian accommodate the case of Sam. In light of this, let me now explain how a Kantian theory of practical rationality can accommodate the case of Sam, given these two claims. Earlier, I noted that the Kantian could not properly accommodate the case of Sam. In order for the Kantian to claim that Sam acts rationally, he must claim that Sam acts in accordance with his best judgment. Moreover in order to say this, he must suggest that Sam has a best judgment telling him he ought to study with friends for reasons ψ. However, the Kantian also suggests that best judgment is the product of deliberation, and
Sam doesn't seem to deliberate in coming to the best judgment that he ought to study with friends for reasons ψ. Moreover Sam doesn't seem to even be aware of this best judgment. Keeping this in mind, how do our new claims help the Kantian explain away the case of Sam?

To see how, I suggest we take a closer look at Sam again. Arpaly suggests that Sam is an ordinary college student who has previously had exams come up in the past. Sam usually deliberates and comes to the best judgment that he ought to study in isolation for reasons φ. Now, as I see it, if Sam really is an ordinary college student, he has a rather complex understanding of the ways in which various work strategies lead to varying degrees of success in his academic pursuits. In fact, it would be rather implausible to suggest that Sam possesses anything less than this rather complex understanding. Of course, Sam's thoughts on studying aren't entirely refined, since he seems to believe that studying in isolation will lead him to great success on his exams. However, he does have a reasonable view of how various work strategies result in various degrees of success. Keeping this in mind, we might think that Sam, at least prior to his eventual action, has a rather sophisticated conception of the good, at least when it comes to studying and exams. Sam values success on his exams, and moreover he also values a variety of different work strategies as a means to obtaining this success. We would also expect that Sam's conception of the good contains comparative value judgments regarding the relative priority of these values as well. So, for instance, Sam values studying in isolation more than he values studying with friends, and he values

96 If Sam doesn't have this rather complex understanding, I submit that Sam probably doesn't have the reasoning skills necessary to get into college in the first place.
studying with friends more than he values not studying at all.

Of course, on previous occasions, Sam has deliberated and formed the best judgment that he ought to isolate himself for reasons \( \phi \), and on these previous occasions, Sam has chosen to act in accordance with this best judgment. So, Sam has a best judgment in his conception of the good that tells him he ought to study in isolation for reasons \( \phi \) (via Thesis 1). Now, Sam comes to believe that extreme isolation never actually gets him success on exams. He never reflects on this belief in any normatively significant way. However, there is something interesting about this belief that is worth noting. The belief that studying in isolation results in success on exams is an “upstream” belief that plays a justifying role in Sam's value of studying in isolation. Moreover, Sam's value of studying in isolation is itself an “upstream” value that plays a justifying role in his comparative value judgment that studying in isolation is more valuable than studying with friends. Also, this comparative value judgment is one that plays a justifying role in his best judgment that he ought to study in isolation for reasons \( \phi \).

Lastly, Sam deliberated from his belief about isolation to his value of studying in isolation, to his comparative value judgment, and finally to his best judgment about studying in isolation.

Given that “downstream” elements of an agent's conception of the good can be unconsciously modified to reflect consistency with “upstream” changes that are made (i.e., Thesis 2), Sam's new belief causes a number of unconscious “downstream” changes in Sam's conception of the good. So, for instance, Sam unconsciously ceases to value studying in isolation. As a result of this, Sam's comparative value judgment that studying
in isolation is *more valuable* than studying with friends is unconsciously modified to become a comparative value judgment that *studying with friends is more valuable*. As a result of this, Sam's best judgment that he ought to isolate himself for reasons $\varphi$ is unconsciously modified to become a *best judgment that he ought to study with friends for reasons $\psi$*. In each of these cases, the “downstream” value is modified to reflect consistency with an “upstream” change that has been made. Since these changes do not require the *addition* of new judgments to the agent's conception of the good, Sam reaches his new best judgment without needing to deliberate on it. Moreover, since these changes occur via *content-efficacious rationalizing cause*, Sam forms this new best judgment without being aware that he has the best judgment in question. This explanation sounds quite complicated, but I think it captures a simple idea – that our views about how we ought to act can change in ways that we don't consciously reflect on.

**Choice and Awareness of the Reasons one Chooses to Act On**

So far, I've explained how Sam comes to have his unconscious best judgment, without deliberation being directly involved. At this point, we can turn to the second problem regarding choice. Specifically, I suggested in the last chapter that a theory of practical rationality cannot make sense of the idea that action is guided by the agent's understanding of the reasons that he commits himself to act on, without incorporating the concept of choice. This suggestion motivated the incorporation of choice into a theory of practical rationality, however, the Kantian takes choice to carry with it an awareness requirement. When one chooses to act for a reason, one is necessarily aware of the
reason that one chooses to act on. This is problematic because in order to say that Sam acts rationally, the theory must say that Sam chooses in accordance with his best judgment. In order to say this, the theory must say that Sam chooses to study with friends for reasons \( \psi \). However Sam is not aware of the reasons that he allegedly commits himself to act on. Keeping this in mind, in order to incorporate choice back into the picture, we have to explain how an agent can choose to act for a reason without thereby being aware of the reason that he commits himself to act on. In order to see a way out, let me argue for two claims. The first is that the choices of an agent are retained by the agent after he chooses, just in the way that best judgments are retained by the agent after he forms them (henceforth “Thesis 3”). The second is that, when an agent chooses to act on the reasons picked out by his best judgment, the agent's choice has a rationalizing relationship with his best judgment (henceforth “Thesis 4”).

Regarding choice, why would I claim that the choices of an agent are retained by the agent after he chooses? Well, let me start out by noting that this claim is supported by some earlier claims I made about the Kantian view. In the second chapter, I noted that Kantians take agents to choose maxims of action. When an agent chooses, what he specifically chooses is a maxim of the form: “perform action X for reasons Y.” The maxim is not one that specifies a particular token action and a particular token reason for action. Rather the maxim specifies a type of action and a type of reason for action. This is important, because if maxims specify types of actions and types of reasons, then it appears that choosing a maxim of action involves a commitment that persists long after one has acted. Choosing a maxim signifies that one commits oneself to performing that
type of action *whenever* that type of reason comes up in the future. This is what I mean when I say that the choices of an agent are retained by the agent after he chooses. This is what Kant means when he speaks of an agent's choosing a maxim as involving *prescribing a rule for oneself*. Keeping this in mind, I think that a look at the content of maxims reveals that the choices of an agent are retained by the agent after he chooses.

There is a surprising passage contained in the Lectures in Ethics which supports this claim as well. Here, Kant connects chosen maxims of actions with *dispositions*.

> “Ethics is thus a philosophy of dispositions, and hence a practical philosophy, for dispositions are basic principles of our actions and serve to couple actions with their motivating ground. It is hard to explain what we understand by a disposition; a person who pays his debts, for example, is not yet on that account an honourable man, if he does it from fear of punishment, etc.; yet he is nonetheless a good citizen, and his action has *rectitudo juridica*, though not *ethica*. But if he does it because of the inner goodness of the action, his disposition is moral and has *rectitudo ethica*.“\(^{97}\)

In this passage, Kant points to the idea that the evaluation of an action requires a look at the chosen principle of action or maxim that underlies the action. That Kant says this is not a surprise. What *is* a surprise is that Kant connects principles of action with *dispositions*. Kant does not at first say that ethics is a philosophy of principles of action, what he says is that ethics is a philosophy of *dispositions*. When Kant speaks of dispositions, I am quite sure he does not mean “disposition” in the sense that I employed in the second chapter – an unconscious disposition to behave in a certain way given a certain set of circumstances. He does speak of dispositions as *being* principles of action or maxims, and given his statements above he seems to speak of

\(^{97}\) LE, 27:299.
dispositions as referring to the *chosen* maxims of the agent. However, I think his use of the term “disposition” also shows that he takes the chosen maxims of the agent to refer to *states* of the agent.\(^98\) When an agent chooses a maxim, he retains a disposition to perform a type of action for a certain type of reason. Keeping this in mind, there is some reason to think that Kant himself thinks that when an agent chooses to act for a reason, his choice is retained long after he acts.

Now, the second claim I would like to make is that, when an agent chooses to act on the reasons picked out by his best judgment, the agent's choice has a *rationalizing relationship* with his best judgment (i.e., Thesis 4) – the agent's best judgment *rationally supports* his choice. Why might I make this claim? Well, first of all, I think it is pretty intuitive – we seem to cite our best judgments in explaining our choices all the time. However, I think that this claim can also be taken from some observations that I made in the second chapter. As I noted, Kant states that:

> “Everything works in accordance with laws. Only a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles, or has a will. Since *reason is required for the derivation of actions from laws*, the will is nothing other than practical reason.” *(my emphasis).*\(^99\)

As I argued in the second chapter, Kant takes the agent's best judgments and choices to have a special psychological relationship with one another. Given the context of the quote, Kant essentially says here that the agent's choices are *derived from* the agent's best judgment by means of the faculty of reason. This is what Kant means when he says that reason is required for the derivation of action from laws. Of course, if the

---

\(^98\) Given Kant's use of the term “disposition” in later works, this is not a stretch.  
\(^99\) G, 4:412.
choices that an agent makes are actually derived from the agent's best judgments, then it appears that Kant does think that, at least in cases where the agent chooses to act on the reasons picked out by his best judgment, the agent's choice has a rationalizing relationship with his best judgment.

**How One Chooses without Awareness of the Reasons one Chooses to Act On**

I bring these two claims up, because I think that they make room for the possibility that an agent can choose to act for a reason without thereby being aware of the reason that he chooses to act on. What I would like to propose is that the mechanism that allows agents to come to best judgments without deliberation is the very same mechanism that allows agents to choose to act for reasons without being aware of the reasons that they commit themselves to act on. Specifically, if the choice of an agent is retained in the form of standing commitment to act in a certain way for a certain type of reason (Thesis 3), and when an agent chooses to act on the reasons picked out by his best judgment, his choice has a rationalizing relationship with his best judgment (Thesis 4), then it seems natural that the mechanism employed earlier with regard to deliberation would also apply to choice. After all, the agent's choice is essentially a “downstream” state in relation to his “upstream” best judgment (as Kant suggest, it is “derived” from his best judgment). If “downstream” states of the agent can be unconsciously modified to reflect consistency with the agent's “upstream” changes, then an agent's choices can change as well given changes to his best judgments. We need only add to this the qualification that the rationalizing relationship that exists between the agent's best
judgment and choice only causes a change in the agent's choice when it is the case that the agent has correctly reasoned from the best judgment to the choice – that is, when the agent has rationally derived the choice in question from the best judgment in question.

So say that an agent deliberates and comes to the best judgment that he ought to X for reasons φ, and say that he acts in accordance with his best judgment by choosing to X for reasons φ. Say moreover that he comes to a new belief that causes unconscious changes in his conception of the good. The belief can cause an unconscious modification of his best judgment, resulting in a best judgment that he ought to Y for reasons ψ. However, given that his retained choice has a rationalizing relationship with his best judgment, and given that his best judgment is “upstream” of his retained choice, we can expect that his retained choice will be unconsciously be modified to become a choice to Y for reasons ψ. Moreover, given that the change takes place unconsciously, it will be the case that he chooses to Y for reasons ψ without being aware of the reasons that he commits himself to act on. Of course, looking at this story at the level of mental states makes the story sound extremely complex. However, I think that there is a simple intuition that it explains – sometimes an agent's commitment to act can be influenced by his views on how he ought to act, even though he is not aware of the influence.

**Analysis of Sam's Case**

So far I have provided a new story about how best judgments are formed, and how choices are made. We now have a Kantian theory of practical rationality that explains how agents can form best judgment without deliberation, and how they can
choose to act for reasons without being aware of the reasons that they choose to act on.

At this point, we can get back to the case of Sam again. I earlier provided the story of how Sam comes to form the best judgment that he ought to study with friends for reasons $\psi$. Sam's belief that studying in isolation doesn't lead to success on exams causes a series of unconscious changes in Sam's conception of the good. One of these changes involves the unconscious modification of Sam's best judgment that he ought to study in isolation for reasons $\phi$ – this best judgment is modified to become a best judgment that he ought to study with friends for reasons $\psi$. What happens after this?

Well, given that Sam previously chose to act on the reasons picked out by his best judgment (the one telling him he ought to study in isolation for reasons $\phi$), Sam retains this choice in the form of a continued commitment to study in isolation for reasons $\phi$ (Thesis 3). Moreover, given that Sam chose to act on the reasons picked out by his best judgment, Sam's choice has a rationalizing relationship with his best judgment (Thesis 4). Given that Sam's retained choice is “downstream” of his best judgment that he ought to study in isolation for reasons $\phi$, and given that his best judgment was unconsciously modified to become a best judgment that he ought to study with friends for reasons $\psi$, Sam's retained choice is also unconsciously modified to maintain consistency with the change in his downstream best judgment. Sam comes to choose to study with friends for reasons $\psi$, and moreover retains this choice in the form of a continued commitment to study with friends for reasons $\psi$. Given that this modification took place unconsciously, Sam was not aware of reason that he committed himself to act on.

There is one last piece of the puzzle to fill in. I have explained how Sam
comes to choose to study with friends for reasons $\psi$ without his being aware of the reason that he commits himself to act on. How is it that he ends up acting? Here, I think we can employ the maxim solution that I suggested in Chapter 2. Given that Sam retains a choice to study with friends for reasons $\psi$, this choice establishes in Sam a *disposition* to study with friends in the case that $\psi$ is present. This disposition in turn is manifested in Sam's action of studying with friends for reasons $\psi$.\footnote{100}{Why bring the maxim solution in here? Well, the belief that causes the chain reaction was presumably acquired long before his eventual action. Given this, the change in his choice took place long before his eventual action as well. Keeping this in mind, there must be something that connects his past choice to his future action. The maxim solution provides this connection.}

Since Sam's action is a manifestation of a disposition to act, Sam doesn't need to be aware of the *reasons on which he acts*. Since Sam's choice is formed via the mechanism I outlined earlier, Sam doesn't ever need to be aware of the *reasons that he commits himself to act on*. Since Sam's best judgment is formed via the mechanism I outlined earlier, Sam doesn't need to be aware of the *reasons he ought to act on*. Nor does he need to deliberate *at any time* in reaching his best judgment. Since Sam acts *in accordance with his best judgment*, Sam acts rationally.\footnote{101}{Again, as before, given that acting in accordance with a best judgment is taken to be a necessary (not sufficient) condition for rational action, and acting against a best judgment is taken to be a sufficient condition for irrational action, it is possible for one to still suggest that Sam acts irrationally. However, in order to suggest this, one would have to both provide a new condition for irrational action and suggest that Sam satisfies this condition. In the absence of such a story, there is overwhelming reason to think that Sam acts rationally, given my analysis.} Moreover, since the reasons that Sam acts on are *not* contained in the domain of his conscious best judgment, Sam doesn't act against his conscious best judgment. As a result, the Kantian is not compelled to say that Sam also acts irrationally.\footnote{102}{We can also add here, under this new story, Sam's choice isn't made *given his conscious best judgment*. In light of this, Sam doesn't act against it and hence doesn't act irrationally.} Since Sam acts on the reasons picked out by his unconscious best judgment, he acts *for the best reasons*. Moreover, since Sam acts on his...
best reasons because his best judgment tells him he ought to act on them, Sam acts on his best reasons because they are the best reasons. Given that this new theory of practical rationality accommodates the case of Sam, and given that it is a Kantian theory of practical rationality, we now have a defense of a Kantian theory of practical rationality. Finally, since this modified Kantian theory of practical rationality has a greater explanatory power than the original Kantian theory of practical rationality outlined in the first chapter, it counts as an improvement over the original Kantian theory of practical rationality.

**Final Objections**

So far, I have presented a modified version of the Kantian theory of practical rationality. At this point, I would like to look at a significant trio of criticisms that one might have for the new theory. The first criticism is that the modifications I have made are unmotivated from the Kantian perspective. The second criticism has to do with the charge that my view over-intellectualizes action. The third criticism has to do with the dilemma that I mentioned at the end of the last chapter. Regarding the first criticism, let me say the following: At least as I see it, all of the changes that I have made to the Kantian theory of practical rationality are either supported by claims that Kantians have made or consistent with claims that Kantians have made. Thesis 1, for instance, is supported by Herman's own characterization of the agent's conception of the good. Thesis 3 is supported by Kant's own statements regarding the content of maxims, and by Kant's characterization of chosen maxims in the *Lectures on Ethics*. Thesis 4 is
supported by Kant's statements on the relationship between choice and best judgment. Thesis 2 is not supported by any particular Kantian claims, however I showed that it was consistent with Herman's conception of the role of deliberation. In light of this, I think that I have good reason to characterize this theory of practical rationality as a specifically Kantian theory of practical rationality.

Regarding the second criticism, there are a few ways to interpret the charge of over-intellectualization. If the criticism is specifically that the theory I have presented is psychologically implausible from a phenomenological standpoint, I believe that the criticism is unwarranted. The criticism might be warranted if I had retained the Kantian claim that best judgments and choices must be conscious. However, given that my theory has it that Sam's best judgments and choices are unconscious, I think I have properly accommodated Sam's action in full. Moreover, given that my account of Sam's psychological history is also plausible, there is no good reason to think that my theory over-intellectualizes Sam's action, at least in this sense. Now, perhaps the criticism is specifically that the theory I have presented is too complicated in that it employs a number of mental states that are unnecessary for accommodating the case of Sam, especially given that Arpaly's much simpler theory accommodates the case of Sam quite well. Regarding this criticism, I grant that my theory of practical rationality is more complicated than Arpaly's. However, if one takes conscious deliberate action to be the paradigm case of agency in action, and one takes agency to be a definitive characteristic of action, then it appears that my theory of practical rationality is not too complicated. The extra complications give the theory an explanatory power that Arpaly's theory does
not have – it accommodates the intuition that action is behavior that is guided by the
agent's understanding of his action. So long as one takes this intuition as demanding
explanation, my theory of practical rationality is not too complicated – rather, Arpaly's
theory of practical rationality is too simple.

We can now turn to the third criticism – one that relates to the dilemma I
presented at the end of the last chapter. In presenting this new theory of practical
rationality, I have essentially claimed that agents can form best judgments without being
aware of the reasons they ought to act on, and can form choices without being aware of
the reasons that they commit themselves to act on. One might find these claims to be
quite un-Kantian in that they directly contradict some of the Kantian motivations for
including best judgment and choice in the first place. At the end of the last chapter, I
argued that one could not properly make sense of the idea that action is guided by the
agent's understanding of the reasons he ought to act on, unless one characterizes best
judgment as necessarily being conscious. I also argued that one could not make sense of
the idea that action is guided by the agent's understanding of the reasons he commits
himself to act on unless one characterizes choice as necessarily being conscious. In both
cases, the agency in action was accounted for by connecting the representations that
reflected the agent's understanding to the awareness of the agent. In light of this, haven't
I failed to construct a theory in which action is properly guided by the agent's
understanding of the reasons he ought to act on, and the reasons that he commits himself
to act on?

Here, let me simply suggest that unconscious best judgments and choices can
be representative of the agent's understanding in action. In other words, the Kantian claim about the connection between consciousness and the agent's understanding is false. I think it is instructive at this point to turn back to the case of Gary and Sally, presented earlier in this chapter. Sally and Gary are examples of agents whose values plausibly change in such a way that does not require conscious deliberation. Gary, specifically, is an agent who modifies one of his value judgments without the modified value judgment being a conscious state of his mind. Now as I see it, it would be quite odd to think that this change in Gary's conception of the good does not correspond to a change in Gary's understanding of how he ought to act in certain situations. In light of this, so long as we take Gary and Sally to be realistic, it appears that Gary and Sally are good evidence that the earlier Kantian claim about consciousness and the understanding is false.

Let me provide some other evidence that unconscious best judgments and choices can be representative of the agent's understanding in action. Note that in the case of Sam, Sam's unconscious best judgment (the one telling him he ought to study with friends for reasons ψ) is one that is created as a result of his deliberation. The rationalizing relationships that directly cause Sam to form his unconscious best judgment can only cause him to form this best judgment in the case that Sam has previously deliberated from certain downstream values and beliefs to the best judgment that he ought to study in isolation for reasons φ. Given moreover that deliberating from these downstream values and beliefs to this best judgment requires that Sam appreciate the rationalizing relationships between his downstream values and beliefs and his best judgment, it appears that Sam's appreciation of the relevant rationalizing relationships is
also a necessary condition for the formation of Sam's unconscious best judgment. This, I think, is good evidence that Sam's unconscious best judgment does in fact represent his understanding of the reasons he ought to act on.

Of course, even given these arguments, there may still be some Kantians who are unhappy with the revisions I have made – though it's not clear to me that most Kantians will find such revisions objectionable. If it is any consolation to these Kantians, Arpaly herself is also likely unhappy with the new theory that I have proposed. Moreover, looking at the bigger picture, the revisions do help the Kantian secure the claim that action is behavior that is guided by a series of representations that reflect the agent's understanding of his action. They also help to secure that other centuries-old saw about rational and irrational action – that rational action is action that accords with one's best judgment, and irrational action is action that goes against one's best judgment.\textsuperscript{103} Given that most Kantians are proponents of these claims, I think the minimal revisions I make are a small price to pay. The fact of the matter is, you can't make everyone completely happy. But if you can make everyone just a little bit unhappy, in the course of preserving ancient truths, perhaps that's good enough.

\textsuperscript{103} Again, acting in accordance with one's best judgment is taken to be a necessary condition of rational action, and acting against one's best judgment is taken to be a sufficient condition of irrational action.
References


